Attorney Docket No. 35718/237251(5718-134) PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re:	Gruis *et al.*	Confirmation No.:	7358
Appl. No.:	09/934,066	Group Art Unit:	1638
Filed:	August 21, 2001	Examiner:	Stuart F. Baum
For:	METHODS OF INCREASING POLYPEPTIDE ACCUMULATION IN PLANTS		

Mail Stop AF
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

RULE 132 DECLARATION
of
Rudolf Jung

Sir:

I, Rudolf Jung, do hereby declare and say as follows:

1. I am skilled in the art of the field of the invention of the above-referenced application. I have a Doctor rerum naturalis (Dr.re.nat) in Biochemistry from Martin-Luther-University Halle-Wittenberg, Germany. Since 1982, I have been engaged in the study of the processing of plant storage proteins. I have been employed by Pioneer Hi-Bred since 1994.

2. I am a co-inventor of the above-referenced application.

3. Working under my supervision, Sarah Yans and Jan Schulze, Research Associates at Pioneer Hi-Bred International produced *Arabidopsis* plants that were genetically modified to reduce the activity of α-vacuolar processing enzyme, β-vacuolar processing enzyme, γ-vacuolar processing enzyme, ε-vacuolar processing enzyme and

three aspartic proteases. These plants were produced by transforming an *Arabidopsis* line containing knock-out mutations in α-VPE, β-VPE, γ-VPE, and ε-VPE (the "vpe-quad mutant"; see Gruis *et al.* (2004) *Plant Cell* 16:270-90) with a gene silencing vector designed to reduce the activity of three different *Arabidopsis* aspartic proteases (the "AP1-2-3 RNAi vector"). The AP1-2-3 RNAi vector contained sequences corresponding to the following fragments of the *Arabidopsis* aspartic protease mRNA sequences:

NCBI Accession Number	Fragment used for gene silencing vector
NM_104909	nucleotides 1377-1614
NM_101062	nucleotides 1341-1631
NM_116684	nucleotides 1234-1461

The AP-1-2-3 RNAi vector also contained an inverted repeat of this sense sequence, and an intron from the maize alcohol dehydrogenase gene (ADH1) in the spacer region between the sense sequence and the antisense sequence. The use of gene silencing vectors containing inverted repeats for the production of interfering RNA was known to those of skill in the art at the time the present application was filed. *See*, for example Stam *et al.* (1997) *Plant J.* 12:63-82, provided for the convenience of the Examiner as Appendix A; and WO 99/32619 (Fire *et al.*), published July 1, 1999, provided for the convenience of the Examiner as Appendix B.

The *Arabidopsis* vpe-quad mutant plants were transformed by the floral dip method with the AP1-2-3 RNAi vector by *Agrobacterium*-mediated transformation as described by Clough and Bent (1998) *Plant J.* 16:735-43. After self-pollination, hemizygous transgenic seedlings underwent selection based on the expression of a selectable marker gene. The integration of the AP1-2-3 RNAi cassette into the plant genome was confirmed by PCR with primer pairs that amplified a fragment of the RNAi cassette and a fragment of the selectable marker gene. Transgenic plants were then allowed to self-pollinate and the genetic transmission of the transgene was confirmed by selection of transgenic seedlings based on the selectable marker gene.

Protein was extracted from segregating single hemizygous and homozygous transgenic and wild type seeds, and analyzed by SDS-PAGE. Approximately 50-75% of

the seeds collected from several independent transgenic events showed reduced processing of the seed albumin (diminished presence of large and small albumin chains and accumulation of albumin pro-protein precursor) consistent with the expected semi-dominant/ dominant action of the AP silencing cassette. Suppression of albumin processing was not observed in single seed transgenic events in control vpe-quad plants that were transformed with a vector lacking the AP1-2-3 RNAi cassette. The alteration in seed protein processing in the plants transformed with the AP-1-2-3 RNAi cassette demonstrates that this cassette reduced the expression of the corresponding *Arabidopsis* proteases.

4. In a second experiment, soybean plants that were genetically modified to reduce the activity of vacuolar processing enzymes were produced. These transgenic plants were produced using a gene construct that I devised. Based on an experimental plan that I suggested and under the supervision of Zhan-Bin Liu, a Research Scientist at Pioneer Hi-Bred International, the following work was performed. Genetically modified plants were produced by transforming soybean with a gene silencing vector, KS217, designed to reduce the activity of five soybean vacuolar processing enzymes. The KS217 vector had a VPE cassette containing sequences corresponding to fragments of the mRNA sequences of the five soybean VPE's shown below:

Soybean VPE	**Nucleotide sequence used for KS217 vector**
VPE1	nucleotides 1-292 of NCBI Accession No. D28876
VPE1b	nucleotides 12-137 and 1428-1678 of NCBI Accession No. AF169019
VPE2	nucleotides 1-544 of SEQ ID NO: 5 of U.S. Patent Application No. 60/529,666 filed December 15, 2003
VPE2b	nucleotides 1181-1694 of SEQ ID NO:7 of U.S. Patent Application No. 60/529,666 filed December 15, 2003
VPE3	nucleotides 1273-1565 of SEQ ID NO: 9 of U.S. Patent Application No. 60/529,666 filed December 15, 2003

The soybean VPE1 and VPE1b sequences are set forth in NCBI Accession Nos. D28876 and AF169019. The soybean VPE2, VPE2b, and VPE3 sequences are described in U.S. Provisional Patent Application No. 60/529,666 filed December 15, 2003. A copy of this patent application is enclosed for the convenience of the Examiner as Appendix C.

The KS217 vector was constructed with a sense sequence upstream of the VPE cassette, and an inverted repeat of this sense sequence downstream of the VPE cassette. The use of gene silencing vectors containing inverted repeats for the production of interfering RNA was well known to those of skill in the art at the time the present application was filed. *See*, for example Stam *et al.* (1997) *Plant J.* 12:63-82 (Appendix A); and WO 99/32619 (Appendix B), cited above.

Soybean embryonic suspension cultures were transformed with the KS217 vector by particle bombardment essentially as described on pages 37-39 of the instant patent application. The embryos were selected based on the expression of a selectable marker gene, and then regenerated into fertile transgenic soybean plants. Protein was extracted from seeds from these plants, and analyzed by SDS-PAGE. More than 50% of the soybean storage protein glycinin in the transformed seeds accumulated as proglycinin precursor, and this phenotype was found to be stable over at least three generations. The alteration in glycinin processing demonstrates that transformation with the KS217 vector successfully reduced the expression of the corresponding soybean VPE's.

5. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Rudolf Jung

08/24/04
Date

The Plant Journal (1997) **12**(1), 63–82

Post-transcriptional silencing of chalcone synthase in *Petunia* by inverted transgene repeats

Maike Stam, Rob de Bruin, Susan Kenter, Renier A.L. van der Hoorn, Rik van Blokland†, Joseph N.M. Mol and Jan M. Kooter*

Department of Molecular Genetics, Institute for Molecular Biological Sciences, BioCentrum Amsterdam, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

Summary

To induce post-transcriptional silencing of flower pigmentation genes by homologous sense transgenes in transgenic petunias, it is not necessary for the transgenes to be highly transcribed. Even promoterless transgenes can induce silencing. Here it is shown that in these cases silencing is mediated by multimeric transgene/T-DNA loci in which the T-DNAs are arranged as inverted repeats (IRs). With the transgene constructs used, monomeric T-DNA loci are unable to confer silencing even though they modulate IR-induced silencing. IRs with the silencing sequences proximal to the centre (IR_c) induce a more severe silencing than IRs with these sequences distal to the centre (IR_n). Somatic reversion of silencing, as observed in a side branch of one of the chalcone synthase (*Chs*) transformants, was associated with a deletion of the IR locus from L1 cells, the meristematic cell layer that expresses the endogenous *Chs* genes in the flower corolla. Taken together, these data indicate that the post-transcriptional silencing mechanism can be activated by inverted transgene repeats. It is also shown that a silent IR *UidA–ChsA* locus silences the expression of a monomeric 35S promoter-driven *UidA–ChsA* transgene only in corollas where the endogenous *Chs* genes are highly transcribed. These results are consistent with a model in which an IR, by virtue of its palindromic sequence organization, is able to promote the production of aberrant RNAs from the endogenous homologs as a result of ectopic pairing.

Introduction

Gene silencing is a common phenomenon in transgenic plants and affects transgenes and endogenous genes (reviewed by Baulcombe and English, 1996; Matzke and Matzke, 1995; Meyer, 1995, 1996; Stam *et al.*, 1997). If the promoter is inactivated, which is often correlated with DNA methylation, transgenes are transcriptionally silenced (Elmayan and Vaucheret, 1996; Meyer *et al.*, 1993; Neuhuber *et al.*, 1994; Park *et al.*, 1996). If RNA is produced but fails to accumulate, transgenes are post-transcriptionally silenced (De Carvalho *et al.*, 1992; Dehio and Schell, 1994; Depicker *et al.*, 1996; Elmayan and Vaucheret, 1996; English *et al.*, 1996; Goodwin *et al.*, 1996; Ingelbrecht *et al.*, 1994; Mueller *et al.*, 1995; Smith *et al.*, 1994). The expression of endogenous genes can also be post-transcriptionally silenced by introduced sense transgenes when these genes are sufficiently homologous to the endogenous counterparts (De Carvalho Niebel *et al.*, 1995; Kunz *et al.*, 1996; Van Blokland *et al.*, 1994).

How is post-transcriptional gene silencing (PTGS) activated? A few studies indicate that excessive production of transgene RNA might be the trigger (De Carvalho *et al.*, 1992; Elmayan and Vaucheret, 1996; Goodwin *et al.*, 1996; Smith *et al.*, 1994). This occurs efficiently when transgenes are transcribed from a strong promoter (Elmayan and Vaucheret, 1996; Jorgensen *et al.*, 1996) or present in high copy numbers (Dorlhac de Borne *et al.* 1994; Palauqui and Vaucheret, 1995). To explain PTGS, it is assumed that a particular RNA can be produced only up to a certain level. Exceeding this threshold level initiates the degradation of these RNAs. This RNA threshold model gained support from studies of viral transgene-mediated virus resistance in plants (De Haan *et al.*, 1992; Dougherty *et al.*, 1994; Goodwin *et al.*, 1996; Lindbo *et al.*, 1993; Smith *et al.*, 1994). Mainly the transformants in which the transgenes were highly transcribed were resistant (Goodwin *et al.*, 1996; Lindbo *et al.*, 1993; Smith *et al.*, 1994).

Resistance to virus infection is explained by assuming that the mechanism that prevents (viral) transgene RNAs from accumulating also prevents the accumulation of the homologous viral RNA. Even post-transcriptionally silenced non-viral transgenes, such as *UidA* or *NptII*, will prevent infection by a chimaeric virus which carries these non-viral sequences as part of the viral genome (English *et al.*, 1996). As RNA viruses replicate in the cytoplasm, these results suggest that the process of RNA degradation is entirely cytoplasmic (Dougherty and Parks, 1995). It has been proposed that this process involves the action of a plant encoded RNA-dependent RNA polymerase (RdRP, Dougherty and Parks, 1995; Lindbo *et al.*, 1993) which uses the transgene transcripts as a template to

Received 16 October 1996; revised 20 January 1997; accepted 21 February 1997.

*For correspondence (fax +31 20 4447137; e-mail: kooter@bio.vu.nl).

†Present address: Department of Genetics, University of Leeds, Leeds LS2 9JT, UK.

synthesize small complementary RNAs (cRNA). These cRNAs are thought to tag homologous RNAs for degradation by dsRNA-specific ribonucleases (Dougherty and Parks, 1995). The possible involvement of cRNAs (antisense RNA) is attractive as it explains the strong sequence specificity of PTGS. In this model, the question as to how RdRP recognizes only the excessively produced RNAs amongst the thousands of others that are produced remains unanswered. Perhaps only particular RNAs or aberrant RNAs are utilized as a template and these may constitute just a small proportion of the total transgene RNA pool (English *et al.*, 1996).

However, PTGS is not always associated with excessively active transgenes, as is shown for transgene-mediated virus resistance (English *et al.*, 1996; Mueller *et al.*, 1995) and for the silencing of endogenous plant genes (Van Blokland *et al.*, 1994). In the latter case, silencing was induced by a T-DNA carrying a promoterless chalcone synthase (*Chs*) transgene which was not detectably transcribed in the transformants. These results suggest excessive production of transgene RNA is not a prerequisite for activation of the PTGS mechanism.

We are studying the post-transcriptional silencing of the pigmentation gene *Chs* in *Petunia hybrida*. This gene is required for the synthesis of anthocyanin pigments in flowers and its silencing results in fully white flowers or flowers with a variegated pigmentation phenotype (Jorgensen *et al.*, 1996; Napoli *et al.*, 1990; Van Blokland *et al.*, 1994; Van der Krol *et al.*, 1990). Expression of the endogenous *Chs* genes in the corolla of these flowers is down-regulated by a post-transcriptional mechanism, as determined by run-on transcription assays, (Van Blokland *et al.*, 1994). The fact that the transgenes do not have to be highly transcribed indicates that, in this case, PTGS is induced in a way that is different from that of the RNA threshold model. Another observation is that just a minority of the primary transformants show silencing. These transformants not only differ in transgene expression levels but also in transgene copy number, and importantly, the way the transgenes are integrated in the genome: as single copies or as repeats. To determine whether PTGS is associated with the presence of a particular transgene locus, we examined the structure of the transgene loci present in several *Chs* sense transformants in which the endogenous homologs are silenced to various degrees. By performing crosses and by genetic and molecular analysis of the progeny, we identified the T-DNA loci that segregated with the silencing phenotype. None of the monomeric T-DNA inserts identified induced silencing. Silencing was only observed in the plants carrying a multimeric T-DNA locus in which the T-DNAs were organized as inverted repeats (IR), and seemed to require transcription of the endogenous gene(s).

Results

Physical mapping and structure of T-DNA loci in Chs *transformants*

The T-DNA(s) of *Agrobacterium tumefaciens* transformed plants may differ in copy number, integrity, and when multiple copies are physically linked, in their relative orientation. They are usually inserted at different chromosomal sites and are sometimes associated with binary vector sequences (Martineau *et al.*, 1994). As it is not known to what extent these factors affect silencing of endogenous genes, it was important to carefully map the T-DNA loci in the *Chs* silenced transformants previously described by Van Blokland *et al.* (1994) and to study their heritability with PTGS. To be able to do this, the *ChsA* transformants PSE6-2, PSE19-3, PSE19-1-4, PSE21-1 and PSE21-6 (Van Blokland *et al.*, 1994) were first back-crossed to untransformed V26 plants as outlined in Figure 1(b). These transformants carry the transgene constructs pSE19, pSE6 or pSE21, of which the physical maps are shown in Figure 1(a).

T-DNA locus of transformant PSE6-2

The T-DNA locus of this transformant was analysed in two progeny plants (W7016-10 and W7017-10, Figures 1b and 2e). Figure 2 (a)–(c) shows a selection of the Southern blot hydridizations. Figure 2(d) shows the constructed physical map of the T-DNA inserts and the position of the various restriction fragments. *Hin*dIII-digested DNA gives rise to fragments of 5.9 kb (G) and 11.5 kb (F) which hybridize to the *UidA* probe (Figure 2a, lanes H), suggesting two T-DNAs. The *NptII* probe detects a single 5 kb fragment (B, panel (b), lanes H), which is expected if the two T-DNAs are linked and arranged as an inverted repeat (IR) centered around the Right T-DNA border (RB) (IR_n). This IR_n structure is consistent with the *Eco*RI digest. The *UidA* (panel (a), lanes E), *NptII* (panel (b), lanes E) and 3'nos probes (panel (c), lanes E) detect the same 15 kb fragment (E), which is actually larger than the expected 12.4 kb in the case of an IR_n. One of the T-DNAs appeared truncated at the left border, as the *UidA* (panel (a), lanes EH) and 3'nos probes (panel (c), lanes EH) detect in an *Eco*RI/*Hin*dIII double-digest the expected 3.7 kb fragment (C), but also a 5.9 kb fragment (G). The 3'nos-hybridizing 5.9 kb band is about 30% less intense than the 3.7 kb fragment, which suggests that the endpoint of the T-DNA is within the nos polyadenylation region of one of the *UidA–ChsA* transgenes. This was confirmed by an *Eco*RI/*Hin*dIII/*Dra*I triple digest which generated a 4.9 kb *UidA*-hybridizing fragment (D, panel (a), lanes EHD). In addition to the described fragments, some faint bands were visible with *Hin*dIII, *Eco*RI/*Hin*dIII and *Eco*RI/*Hin*dIII/*Dra*I digests (all three panels). The sizes

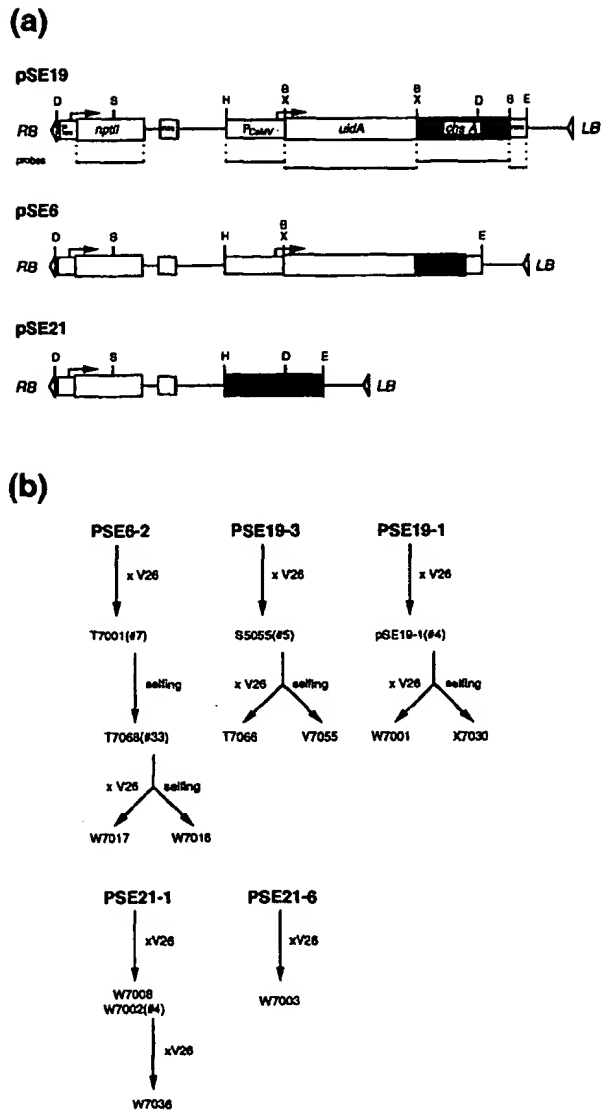


Figure 1 Schematic representation of the *Chs* transgene constructs and overview of the crosses involving the transformants.
(a) Physical maps of the T-DNA constructs used to generate transgenic petunias (Van Blokland *et al.*, 1994). In addition to the selectable marker gene *NptII*, pSE19 and pSE6 contain a chimaeric gene consisting of the *UidA*-coding region fused to the full-length *ChsA* cDNA or the 5' half, respectively; pSE21 contains just the full-length *ChsA* cDNA without a promoter in front of it. Arrows mark the transcription start sites of the nopaline synthase promoter (P_{nos}) or the CaMV 35S promoter. Fragments used as probes for the Southern blot analysis are indicated beneath pSE19 as bars: *NptII*, nos polyadenylation region, P_{CaMV} *UidA*, *ChsA*. Abbreviations: B, *Bam*HI; D, *Dra*I; E, *Eco*RI; H, *Hin*dIII; *LB*, left T-DNA border; *RB*, right T-DNA border; S, *Sph*I; X, *Xba*I; nos, nos polyadenylation region.
(b) Crossing schemes showing the transformants generated by Van Blokland *et al.* (1994) in bold. The progenies are indicated by non-bold letters. The numbers of the plants used for subsequent crosses are indicated in brackets.

of these fragments correspond to partially digested fragments.

T-DNA loci of transformant PSE19-3

The T-DNA inserts of this transformant were analysed in three progeny plants (S5055-8, 2, and 14, Figure 3e) of a back-cross of PSE19-3 to V26 (Figure 1b). Lanes E of Figure 3 show that the *UidA* (panel (a)) and *Chs* probes (panel (b)) detect *Eco*RI fragments of 13.6 kb (C) and 8.2 kb (F). These fragments segregate in a Mendelian manner indicating that they are derived from separate T-DNA loci located on different chromosomes. Fragment C can be derived from a locus that consists of two T-DNAs arranged as inverted repeats with the *NptII* genes near the centre of the IR (IR_n). It has the expected size (13.6 kb) for an IR_n fragment and the hybridization signal is twice as high as that of the other locus, which consists of a single T-DNA (see below). This IR_n structure is consistent with the *Hin*dIII digest, as it generates a single *NptII*-hybridizing fragment of the expected 5 kb (A, panel (c)). The *UidA* (panel (a)) and *ChsA* (panel (b)) probes both detect two *Hin*dIII border fragments, of 4.6 kb (D) and 5.5 kb (E). The *UidA–ChsA* transgenes of the two T-DNAs are intact as the *Eco*RI/*Hin*dIII digest gives the expected 4.3 kb fragment B with the *UidA* and *ChsA* probes (lanes EH of panels (a) and (b), plant no. 2).

The second locus consists of a single truncated T-DNA (S_t). The *UidA* and *ChsA* probes detect one 7.5 kb *Hin*dIII fragment (G, panels (a) and (b) respectively, plant no. 8). Although the *Eco*RI/*Hin*dIII digest shows that the *UidA–ChsA* transgene is intact (4.3 kb, fragment B, panels (a) and (b)), no hybridization was found with the *NptII* probe (plant no. 8, panel (c)). This result together with the detection of just fragment B with the 3'nos probe (not shown) indicates that the S_t locus lacks the entire *NptII* gene. Furthermore, S_t contains pBin19 vector sequences at both sides. The *UidA*-hybridizing *Hin*dIII fragment G and the *Eco*RI fragment F are also detected by a 2.7 kb *Eco*RV pBin19 Left T-DNA Border (LB) probe (not shown). The precise length of these pBin19 sequences has not been determined, but is less than 3.9 kb.

T-DNA loci of transformant PSE19-1

The inserts of this transformant were examined in the progeny of a back-cross of PSE19-1-4 with V26 (Figures 1b and 4e, W7001 progeny). *Hin*dIII generates a 9 kb (M) and a 14 kb fragment (G) which hybridize to the *UidA* (Figure 4a, lanes H), P_{CaMV} (Figure 4b), lanes H) and *ChsA* probes (not shown). These two fragments are from separate loci located on different chromosomes as they segregate in a Mendelian manner (panels (a)–(c), plant nos. 12 and 18 versus 16 and 19). Fragment G can be derived from an IR

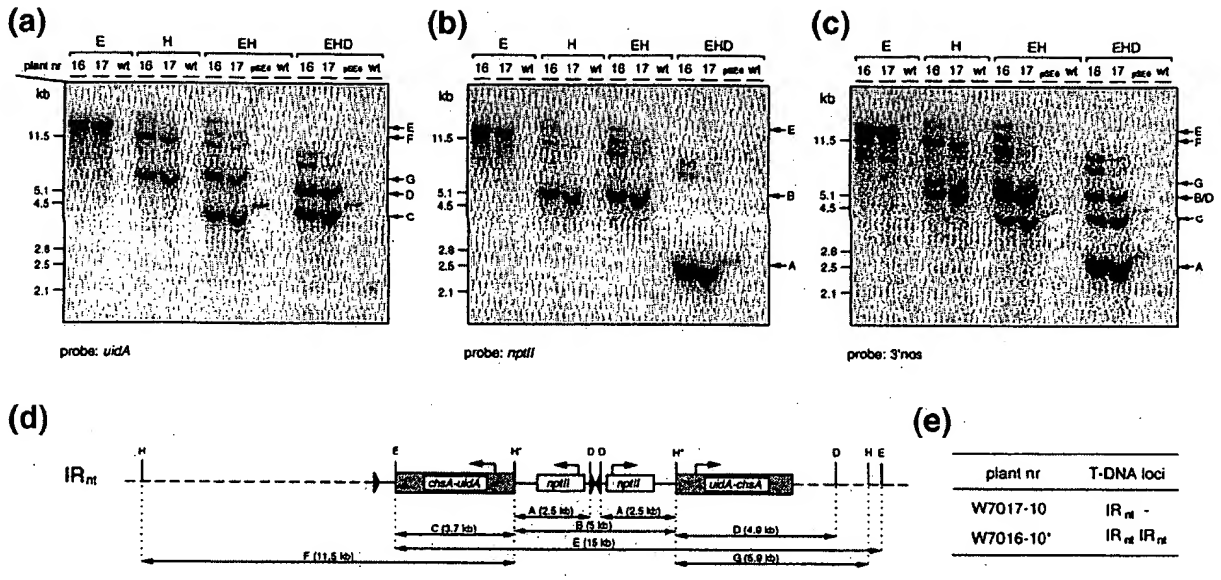


Figure 2. T-DNA locus of transformant PSE6-2.

(a) to (c) show the Southern blot analysis of progeny plants (W7016 and W7017) of PSE6-2 (see Figure 1b). W7016-10 is homozygous and W7017-10 is hemizygous for the T-DNA. Genomic DNA from W7016-10 and W7017-10 digested with *Eco*RI (E at the top), *Hin*dIII (H), both enzymes (EH) or both enzymes with *Dra*I (EHD) was hybridized with a *UidA* probe (a), *NptII* probe (b) or 3'nos probe (c). As a control, pSE6 plasmid DNA was digested with EH and EHD. Lanes indicated by wt contain DNA from untransformed V26 plants. *Pst*I-digested phage lambda DNA was used as a size marker. Capital letters at the right of each panel refer to the fragments in the physical maps in (d).

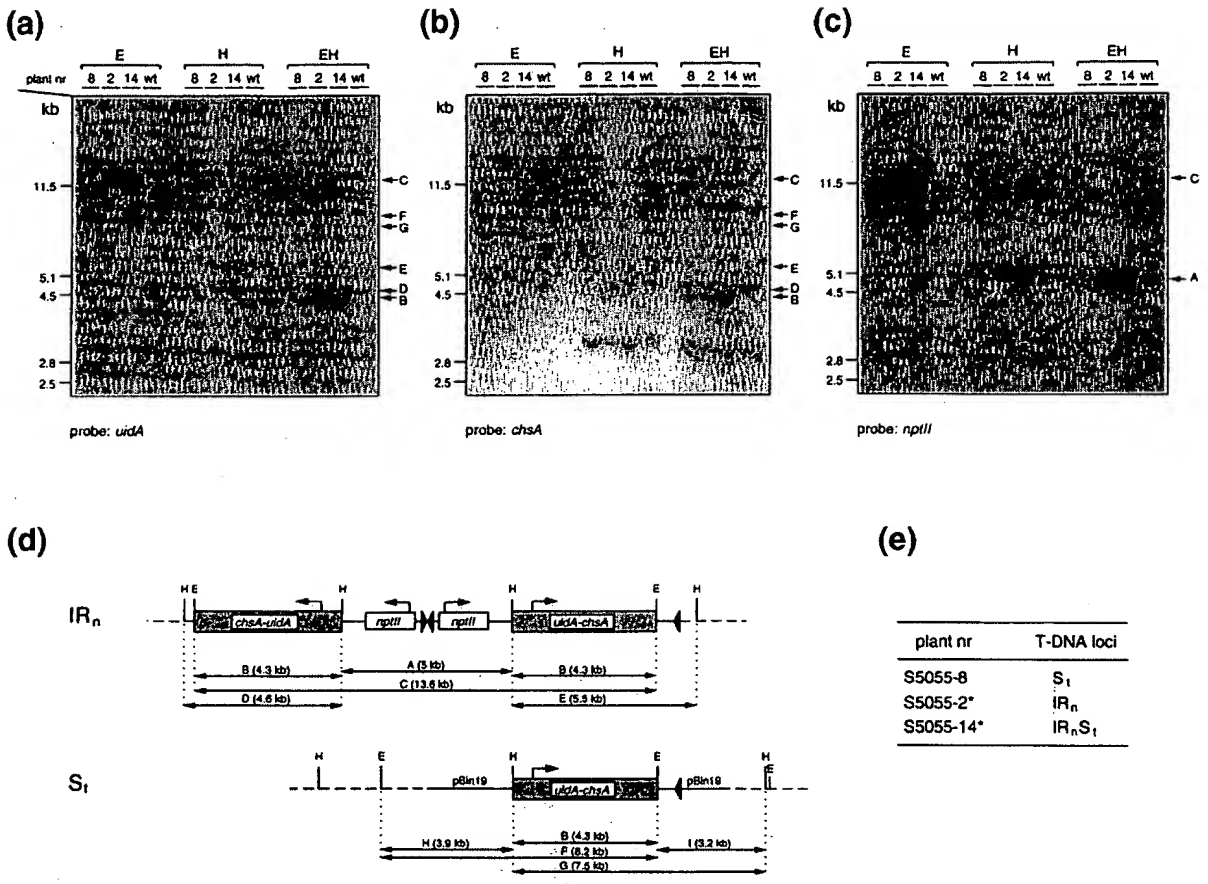
(d) Physical map of the T-DNA locus in W7016-10 and W7017-10 for which the most relevant restriction sites are shown in capitals. The labelling of the fragments refers to the bands on the Southern blots shown in the panels (a), (b) and (c). The interrupted lines indicate flanking plant DNA. The IR_{nt} locus consists of two T-DNAs arranged as an inverted repeat, one of which is truncated at the LB and lacks part of the 3'nos polyadenylation region. H*, a partially-modified *Hin*dIII site.

(e) Summary of the T-DNA loci in the plants shown in (a)–(c). Those indicated by an asterisk produce flowers in which *Chs* expression is silenced.

locus with the *Chs* transgene sequences near the centre (IR_c). While *Hin*dIII gives one fragment G (panel (a), lanes H, plant nos. 16 and 29), *Eco*RI (lanes E, panel (a)) generates two co-segregating fragments of 4.7 kb (F) and 12.5 kb (I) detected by a *UidA* probe. The IR_c structure is consistent with the results of the *Dra*I and *Xba*I digests, which give rise to 2.2 kb (A) and 3.9 kb fragments (D, panel (d)) respectively, both detected by a *ChsA* and 3'nos probe (not shown). However, the A and D fragments are approximately 600 bp smaller than expected for a perfect IR centred around the LB. This suggests that 600 bp between the two adjacent *UidA–ChsA* transgenes are missing, and the locus is therefore termed IR_{ct}. In addition to this truncation, one of the T-DNAs also lacks the *NptII* gene, the CaMV 35S promoter and part of the *UidA* sequence. The 4.7 kb *Eco*RI fragment (F) was not detected by the P_{CaMV} (lanes E, panel (b)) and *NptII* probes (panel (c)), which is consistent with the detection of a single *NptII*-hybridizing *Hin*dIII fragment (C) in plants harbouring just the IR_{ct} (panel (c), lanes H, plant nos. 16 and 29). Quantification of the band intensities by a Phosphor-Imager indicates that about 30% of the *UidA*-coding region is missing. This truncation could be confirmed by an *Eco*RI/*Dra*I double digest which gave rise to a 2.5 kb fragment that hybridized to the *UidA* and *ChsA* probes. For an intact *UidA*-coding region, this fragment should have been at least 2.7 kb.

The DNAs from W7001-29 and PSE19-1-4, digested with *Eco*RI and hybridized with the *UidA*, P_{CaMV} and *NptII* probes (Figure 4a–c) contain a faint 8.2 kb fragment (H) which segregates with the IR_{ct} locus. As the intensity of band H increases, that of band I decreases. We therefore infer that fragment I contains an *Eco*RI site that is partially cleavable (indicated in panel (d) by E*), probably as a result of DNA modification. This modification seems to increase in the successive generations, since fragment H is clearly detectable in PSE19-1-4 (and other plants of that generation) and barely detectable in most W7001 plants.

The second locus appears to consist of a single T-DNA. The *NptII* probe detects one *Hin*dIII fragment of 6 kb (K, panel (c), plant nos. 12 and 18). The *UidA* (panel (a)), P_{CaMV} (panel (b)), *NptII* (panel (c)) and *ChsA* (not shown) probes detect a single 11.5 kb *Eco*RI fragment (L, lanes E). This single T-DNA (S_t) is truncated at the RB. This was concluded from *Eco*RI/*Dra*I and *Eco*RI/*Dra*I/*Hin*dIII digests (data not shown), which showed that the 5' *Dra*I site at position –290 in the P_{nos} promoter of the *NptII* gene is missing. Instead of the expected 2.5 kb *Dra*I–*Hin*dIII fragment, a 4.9 kb fragment (J) was detected by the *NptII* probe. The



plant nr	T-DNA loci
S5055-8	S_t
S5055-2*	IR_n
S5055-14*	IR_nS_t

Figure 3. T-DNA loci of transformant PSE19-3.
(a) to (c) show the Southern blot analysis of progeny plants (S5055) of a back-cross of PSE19-3 to V26. DNA from transformant S5055-8, S5055-2 and S5055-14 digested with *Eco*RI (E at the top), *Hin*dIII (H) or both enzymes (EH) was hybridized with a *UidA* probe (a), *ChsA* probe (b), or *NptII* probe (c). Bands in panel (b) not indicated by capital letters are derived from the endogenous *ChsA* genes. See legend of Figure 2 for further details.
(d) Physical maps of the T-DNA inserts in S5055. The S_t locus in plant 8 consists of a truncated T-DNA lacking the *NptII* gene and is at both sides flanked by pBin19 vector sequences. The IR_n locus of S5055-2 consists of two complete T-DNAs arranged as an inverted repeat with the *NptII* genes in the middle. S5055-14 contains both loci.
(e) Summary of the T-DNA inserts in the S5055 plants shown in (a)–(c). Those indicated by an asterisk produce flowers in which the expression of *Chs* is silenced.

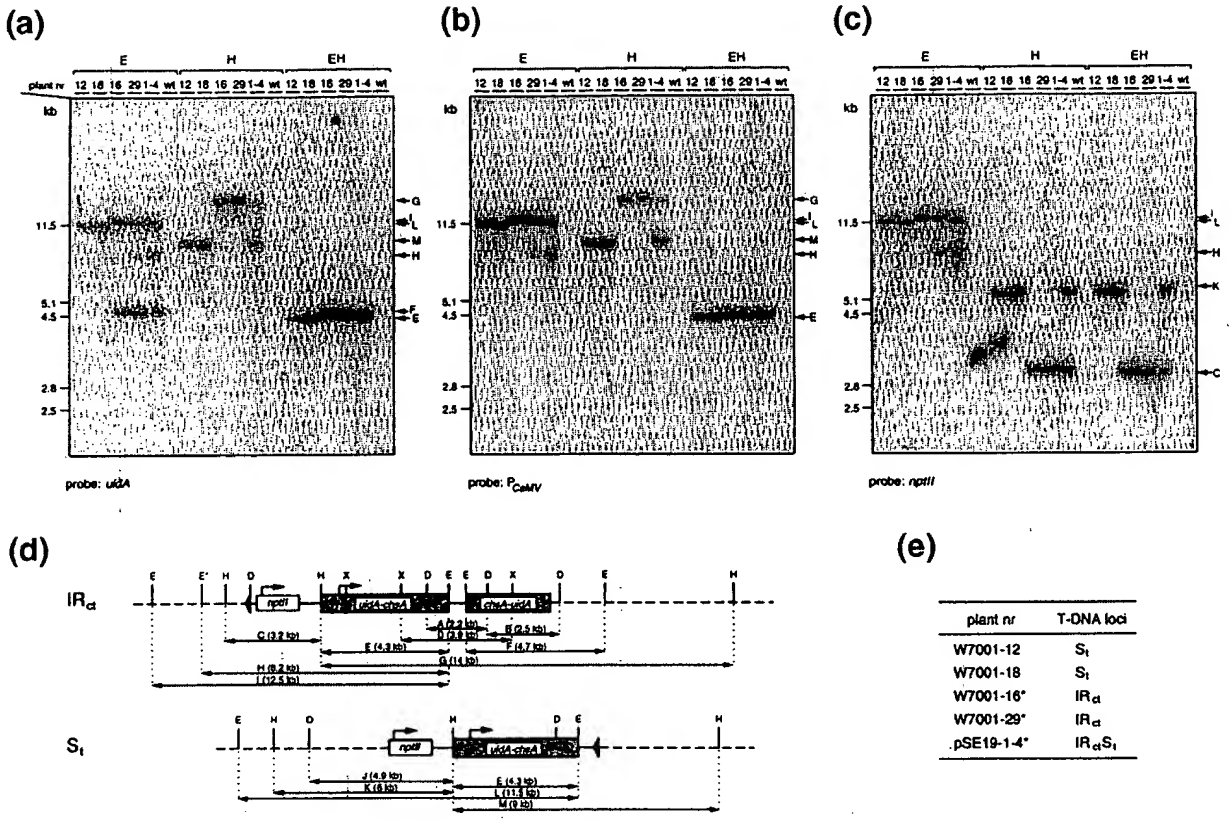
exact breakpoint was not determined but since a normal *NptII* mRNA is produced (not shown), the *NptII* coding region is intact, as well as part of the nos promoter.

T-DNA loci of transformant PSE21-1

The T-DNA inserts of this transformant were examined in four progeny plants of a back-cross of PSE21-1 to V26 (W7002 progeny, Figures 1b and 5d). *Hin*dIII generates fragments of 4.2 kb (B), 5 kb (C) and 3 kb (L) that hybridize to the *NptII* probe (Figure 5a, lanes H). The fragments B and C co-segregate and are derived from a T-DNA locus harbouring three T-DNAs which are all inverted relative to one another (IR_{cn}). Fragment C fits with an IR fragment which contains two *NptII* genes which are centred around the RB. In addition to the endogenous *ChsA* gene fragments, the *ChsA* probe (panel (b), lanes H) detected two *Hin*dIII fragments of 4.2 kb (D) and 3.2 kb (E). Fragment D has the expected size of an IR-fragment centred around the LB and which carries two *ChsA* transgenes. The proposed IR_{cn} structure is consistent with the *Eco*RI digest as the *NptII* (panel (a), lanes E) and *ChsA* probes (panel (b), lanes E) detect a 5.3 kb fragment (F) and a fragment of the expected 8 kb (G). The double IR configuration was confirmed by *Dra*I and *Sph*I digests (H and I, panel (c)).

The second locus consists of a monomeric T-DNA (S). Consistent with this organization is that the *ChsA* probe detects a single 7.5 kb *Hin*dIII fragment (M, panel (b), lanes H) and *Eco*RI gives rise to a single 6.3 kb fragment (N) detected by the *NptII* (panel (a)) and *ChsA* probes (panel (b)).

The T-DNAs of both loci are intact, as the *Eco*RI/*Hin*dIII (panels (a) and (b), lanes EH), *Eco*RI/*Dra*I (not shown) and *Eco*RI/*Hin*dIII/*Dra*I digests (not shown) give rise to the expected fragments with the *ChsA* (panel (b)) and *NptII* probes.



plant nr	T-DNA loci
W7001-12	S_t
W7001-18	S_t
W7001-16*	IR_{ct}
W7001-29*	IR_{ct}
pSE19-1-4*	$IR_{ct}S_t$

Figure 4. T-DNA loci of transformant PSE19-1-4.
(a) to (c) show the Southern blot analysis of progeny plants (W7001) of a back-cross of PSE19-1-4 to a V26 plant. DNA of the plants W7001-12, -18, -16, -29, and the parental plant PSE19-1-4 was digested with *Eco*RI (E), *Hin*dIII (H) or both enzymes (EH) after which the filter was hybridized with a *UidA* probe (a), a P_{CaMV} probe (b), or an *NptII* probe (c). See legend of Figure 2 for further details.
(d) Physical maps of the T-DNA loci in W7001. The S_t locus in the plants 12 and 18 consists of a single T-DNA truncated at the RB. The locus of the plants 16 and 29 consists of two T-DNAs arranged as an inverted repeat (IR_{ct}) with the *UidA–ChsA* genes near the centre. One of the T-DNAs is truncated at the RB, lacking the *NptII* gene, the P_{CaMV} and part of the *UidA* coding region. The parental plant PSE19-1-4 contains both loci. E*, (partially) modified *Eco*RI site.
(e) Summary of the T-DNA loci in the W7001 plants shown in (a)–(c). Those indicated by an asterisk produce flowers in which *Chs* expression is silenced.

Twenty-eight plants contained both T-DNA loci, five plants contained just the IR_{cn} locus. No plants were obtained with only the S locus. These results suggest that PSE21-1 harbours two T-DNA loci on the same chromosome. A χ^2 test indicated that the IR_{cn} and S loci are separated by at least 16 cM.

T-DNA loci of transformant PSE21-6

The inserts of this transformant were examined in six progeny plants of a back-cross of PSE21-6 to V26 (W7003 progeny, Figures 1b and 6f). In addition to the fragments derived from the endogenous *ChsA* genes, *Hin*dIII generates fragments of 4.2 kb (D, plant nos. 14 and 31), 5.4 kb (H, plant nos. 19 and 67) and 13 kb (K, plant nos. 53 and 62) with the *ChsA* probe (Figure 6b, lanes H). In the progeny, these three fragments segregate in a Mendelian manner indicating that they are derived from three separate loci located on different chromosomes. Fragment D is derived from an IR locus composed of two T-DNAs with the *Chs* transgene sequences near the centre of the IR (IR_c). It has the expected size for an IR_c fragment and the hybridization signal is twice as high as that of the single-copy fragment H (see below). Furthermore, the *NptII* probe (panel (a), lanes H) detects two *Hin*dIII fragments, of 5.6 kb (B) and 3.7 kb (C). The IR_c structure is also consistent with the *Eco*RI digest, as it generated two fragments of 8.8 kb (E) and 7.4 kb (F) detected by the *NptII* (panel (a), lanes E) and *ChsA* probes (panel (b), lanes E).

The 5.4 kb *Hin*dIII fragment (H) is derived from a monomeric T-DNA locus (S), from which the single 2.7 kb *NptII*-hybridizing fragment (G, lanes H) is also derived. *Eco*RI generates a single 5.1 kb fragment (I) detected by the *NptII* (panel (a), lanes E) and *ChsA* probes (panel (b), lanes E).

The T-DNAs of the IR and the S locus are intact as the *Eco*RI/*Hin*dIII (panels (a) and (b), lanes EH) and *Eco*RI/*Hin*dIII/*Dra*I (not shown) digests produce the expected fragments hybridizing to the *ChsA* (panel (b), lanes EH) and *NptII* probes (not shown).

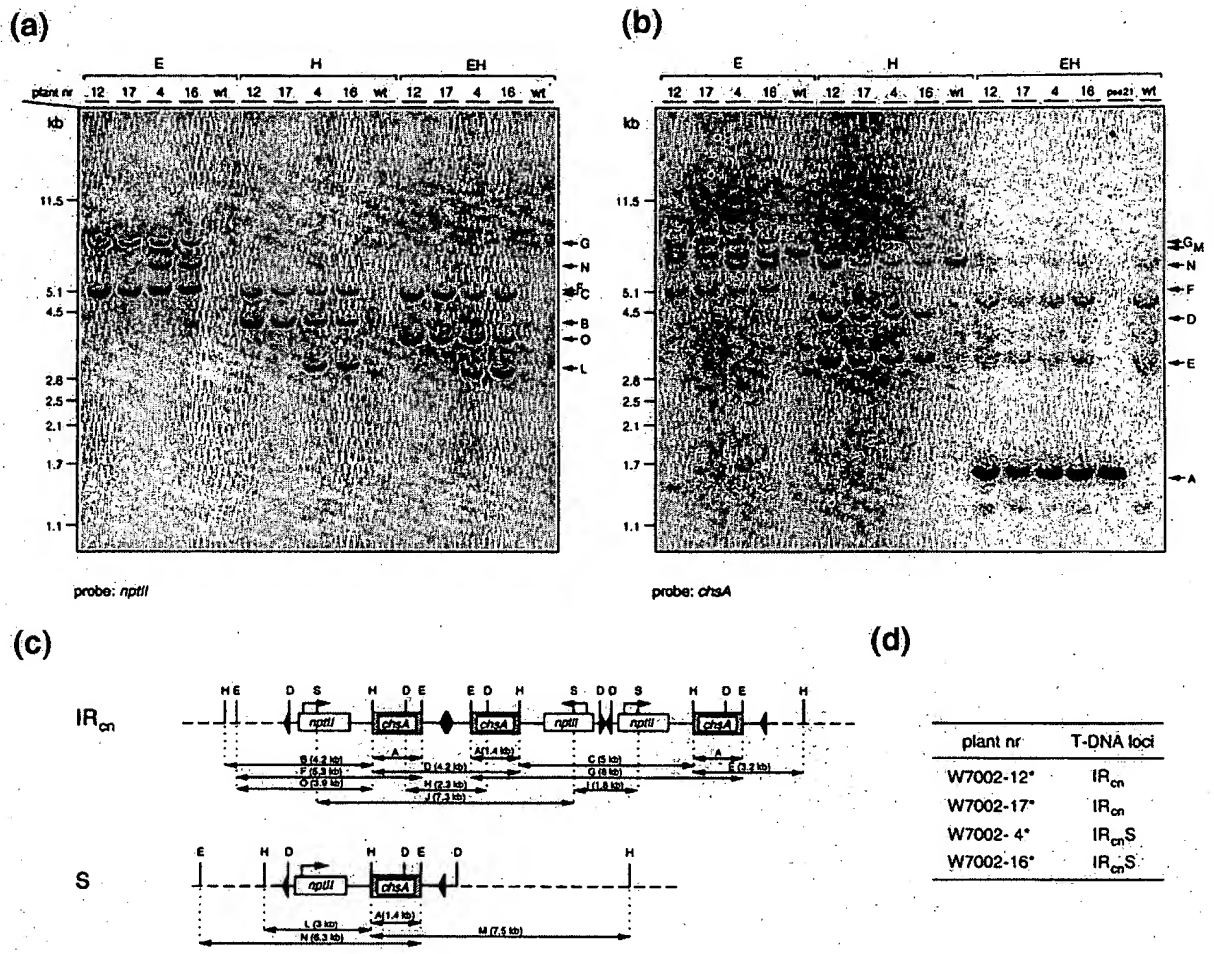


Figure 5. T-DNA loci of transformant PSE21-1.

(a) and (b) show the Southern blot analysis of progeny plants (W7002) of a back-cross of PSE21-1 to V26. DNA from transformant W7002-12, -17, -4 and -16 was digested with *EcoRI* (E), *HindIII* (H) or both enzymes (EH) and the Southern blot filter hybridized with probes for *NptII* (a) and *ChsA* (b). As a control, pSE21 plasmid DNA was digested with EH. Due to a poor transfer of the larger DNA fragments, the repeat-containing fragments C, D and G do not have a higher intensity than the single-copy gene fragments (e.g. F, E, N). However, this was found on other blots (not shown). See legends of Figures 2 and 3 for further details.

(c) Physical maps of the T-DNA loci in W7002. The IR$_{cn}$ locus of the plants 12 and 17 consists of three T-DNAs arranged as inverted repeats. The S locus, additionally present in the plants 4 and 16, consists of a single copy T-DNA.

(d) Summary of the T-DNA loci in the W7002 plants shown in (a) and (b). The asterisk indicates the plants in which the expression of *Chs* was suppressed.

The third locus of PSE21-6 was more difficult to map. The Southern blot data are consistent with a locus comprising one intact and two truncated T-DNAs arranged as direct repeats and separated by the complete pBin19 vector. This locus is called DR'$_{3(2t)}$. The map (Figure 6e) is based on the following observations. *Hin*dIII generates a band of 13 kb which is detected by the *ChsA* probe, but also by the *NptII* and 3'nos probes. Since the *Hin*dIII site in pSE21 (Figure 1a) separates the *NptII* and *ChsA* sequences, this result was unexpected. It can be explained by assuming that the 13 kb fragment consists of two partial T-DNAs arranged in tandem but separated by about 8.4 kb of non-T-DNA. This view is consistent with the *Eco*RI digest which also generates a 13 kb fragment (K') recognized by these probes. Since 8.4 kb is about the size of the pBin19 vector without the T-DNA (8.6 kb, Frisch *et al.*, 1995), the T-DNAs could be separated by pBin19. This was tested by probing the blots with pBin19 probes. Indeed the 13 kb *Hin*dIII fragment K and the 13 kb (K') and 14.5 kb (M) *Eco*RI fragments (panel d) hybridized. The intensity of the *Hin*dIII 13 kb 3'nos-hybridizing band (panel c) was twice as high as that of a single-copy fragment (3 kb, fragment J). This suggested the presence of two identical 13 kb fragments. These fragments are linked because there are just two *Eco*RI plant DNA/T-DNA 3'nos-hybridizing border fragments, of 9 kb (L) (panel c) and 14.5 kb (M). Fragment L is detected by *ChsA* and not by *NptII*, whereas in the case of fragment M it is the other way around, indicating that two of the three T-DNAs in the locus are truncated: one is missing the *NptII* coding region, but still contains the 3'nos

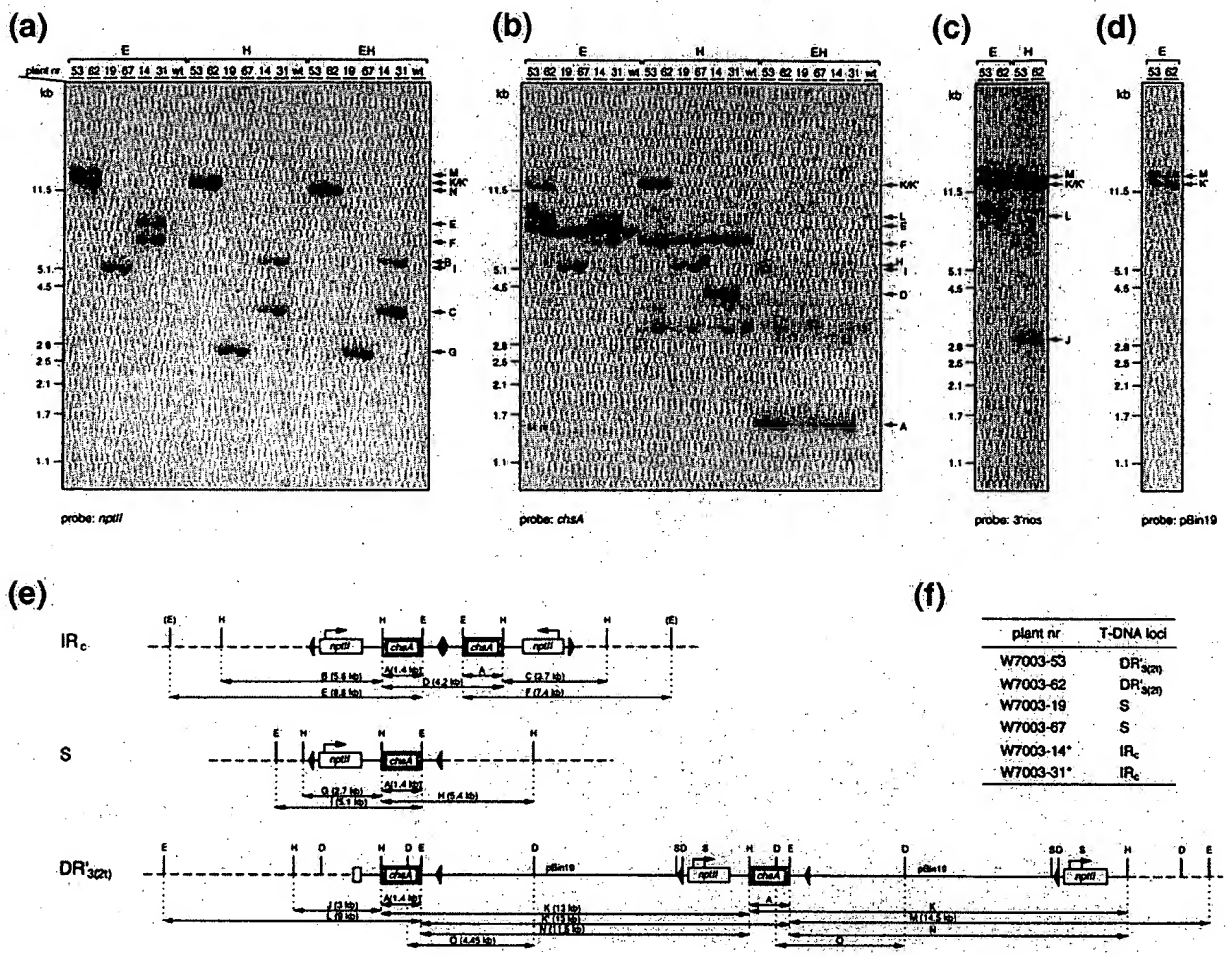


Figure 6. T-DNA loci of transformant PSE21-6.

(a) to (d) show the Southern blot analysis of progeny plants (W7003) of a back-cross of PSE21-6 to V26. Genomic DNA of the plants W7003-53, -62, -19, -67, -14 and -31 was digested with *Eco*RI (E), *Hin*dIII (H) or both enzymes (EH) and the filter hybridized with probes for *NptII* (a), *ChsA* (b), 3'nos (c) and pBin19 (d). The pBin19 probe was a mixture of fragments covering the 8.6 kb pBin19 vector DNA between the RB and the LB. See legend of Figures 2 and 3 for further details.

(e) Physical maps of the T-DNA loci in W7003. The S locus in plants 19 and 67 consists of a monomeric T-DNA insert. The DR'$_{3(2t)}$ locus in plants 53 and 62 consists of three T-DNAs linked in tandem as a direct repeat but separated by pBin19 vector DNA. The T-DNAs at the borders are truncated: one is missing the *NptII* coding region, the other the *ChsA* cDNA. The locus in plants 14 and 31 consists of two intact T-DNAs arranged in an inverted repeat (IR$_c$). (E), the orientation of the *Eco*RI sites with respect to the *Hin*dIII sites is unknown.

(f) Summary of the T-DNA loci in the W7003 plants shown in (a)–(d). Plants marked by an asterisk produce flowers in which *Chs* expression is suppressed.

region, whereas the other is missing the *ChsA* cDNA with the breakpoint just downstream of the *Hin*dIII site. The organization of the T-DNAs was confirmed by various other digests, such as *Dra*I, *Sph*I and *Eco*RI/*Dra*I double and *Eco*RI/*Hin*dIII/*Dra*I triple digests (Figure 6e). Figure 7 depicts the structures of all T-DNA loci detected in the various transformants.

Inheritance of Chs *silencing with inverted T-DNA repeats*

The transformants used for characterizing individual T-DNA loci were also used to examine the inheritance of silencing with these loci. In some instances, transformants were self-fertilized to study the effect of transgene dosage (Figure 1b). Testing seedlings for kanamycin resistance was not useful to follow the segregation of the T-DNA loci as in some instances *NptII* genes were silenced or even deleted from the T-DNA, and more importantly it would not reveal which T-DNA locus was present if the transformant contained two or more loci. The progeny of all crosses was therefore analysed by Southern blotting. To determine whether a plant was homozygous or hemizygous for a particular T-DNA locus, the intensities of the bands on the Southern blots were compared with those of the endogenous single-copy genes chalcone flavanone isomerase (*Chi*) or flavonol synthase (*Fls*). The reliability of this method was verified by PCR analysis on the progeny of back-crosses to untransformed V26 (see Experimental procedures for details).

Figure 7. Summary of the T-DNA loci in the different transformants examined. IR_n, inverted repeat with the *NptII* genes near the centre; IR_c, inverted repeat with the *ChsA* genes near the centre; S, a single-copy T-DNA; DR, a direct repeat; subscript t, T-DNA locus contains a truncated T-DNA. The nos polyadenylation region of the *NptII* gene is indicated by a black dot. Arrows indicate the position of transcription initiation for the transgenes preceded by a promoter.

The segregation data are presented as matrices (Figures 8 and 9) indicating the genotypes (T-DNA locus) and the range of flower pigmentation/silencing phenotypes in the progeny. Each plant is represented by a horizontal bar; its length indicates the degree and variability of *Chs* silencing among the flowers. At least 25 flowers of each plant were used to determine the degree of silencing, which was based on the size of white sectors.

PSE6-2. PSE6-2, which was derived from a selfing (Van Blokland *et al.*, 1994), carries white flowers with purple sectors at the tip of the limbs (Figure 8a). It contains a T-DNA locus that is composed of two T-DNAs arranged as an inverted repeat with the *NptII* genes near the centre: one of the T-DNAs is truncated at the LB and lacks part of the 3'nos polyadenylation region (IR_{nt}; Figures 7 and 2). The progeny (12 plants, T7001) of a back-cross of PSE6-2 to wild-type V26 all contained the IR_{nt} locus and all produced wild-type flowers (not shown). Apparently, PSE6-2 was homozygous for the IR_{nt} locus and this locus only confers silencing in homozygous plants. This was confirmed by examining the progeny (T7068) of a self-fertilization of a hemizygote (T7001-7). Figure 8(a) shows that hemizygous plants do indeed produce wild-type flowers whereas homozygotes produce flowers containing white areas, indicative of *Chs* silencing. The degree of silencing varies from white edges up to almost completely white flowers. Similar results were obtained with the progeny (W7017) of a back-cross of T7068-33, a plant homozygous for the IR_{nt}, to V26; and with the progeny (W7016) of a selfing of T7068-33 (Figure 8a, lower part). Again, silencing only occurred in $IR_{nt}IR_{nt}$ homozygous plants. Note that the degree of silencing in this second series of homozygous plants is reduced compared with that of the first series (T7068).

PSE19-3. Silencing of *Chs* in corollas of PSE19-3 is confined to the region near the tube (Figure 8b, Van Blokland *et al.*, 1994). PSE19-3 is hemizygous for two T-DNA loci (Figures 7 and 3), which are located on different chromosomes. One consists of a single truncated T-DNA lacking the *NptII* gene (S_t) and flanked by pBin19 binary vector DNA. The second locus contains two T-DNAs arranged as an inverted repeat with the *NptII* genes near the centre (IR_n). The inheritance of silencing with these loci was examined in the progeny (T7066) of a back-cross of S5055-5, which is a descendant of PSE19-3 and hemizygous for both loci (IR_n–/S_t–), to V26. Sixty progeny plants were analysed. In plants without a T-DNA locus and in plants with just the S_t locus, *Chs* expression was normal. Silencing was only observed in plants containing the IR_n locus. The silencing phenotype varied from a few white spots near the tube to a clear white ring. To study the effect of IR_n homozygosity, an IR_nS_t plant (S5055-5) was self-fertilized and the progeny examined for the T-DNA loci they contained and their silencing phenotypes. As shown in the lower part of Figure 8b, plants without a T-DNA did not show silencing. This indicates that silencing of the endogenous *Chs* genes is released after the silencing locus is crossed out. Silencing only occurred in plants carrying the IR_n locus. The homozygotes showed a more severe silencing than the hemizygotes. Note that although the S_t locus by itself does not silence *Chs*, not even in homozygous plants (––/S_tS_t), its presence is associated with enhanced silencing by the IR_n locus (IR–/S_t– plants, see also below).

PSE19-1-4. PSE19-1-4 (Figures 7 and 4) which was derived from a back-cross of PSE19-1 to V26 (Van Blokland *et al.*, 1994) contains a monomeric truncated T-DNA insert (S_t) and a two-copy IR locus with the *UidA–ChsA* transgenes near the centre (IR_{ct}). The corollas of PSE19-1-4 were almost completely white with an erratic distribution of purple cells.

The progeny of a back-cross of PSE19-1-4 to V26 (W7001, 55 plants) was analysed for the T-DNA inserts and *Chs* silencing phenotypes. The results are summarized in the upper part of Figure 8(c). Only plants containing the IR_{ct} locus produced flowers with a *Chs* silencing phenotype which varied from a few white spots to almost fully white

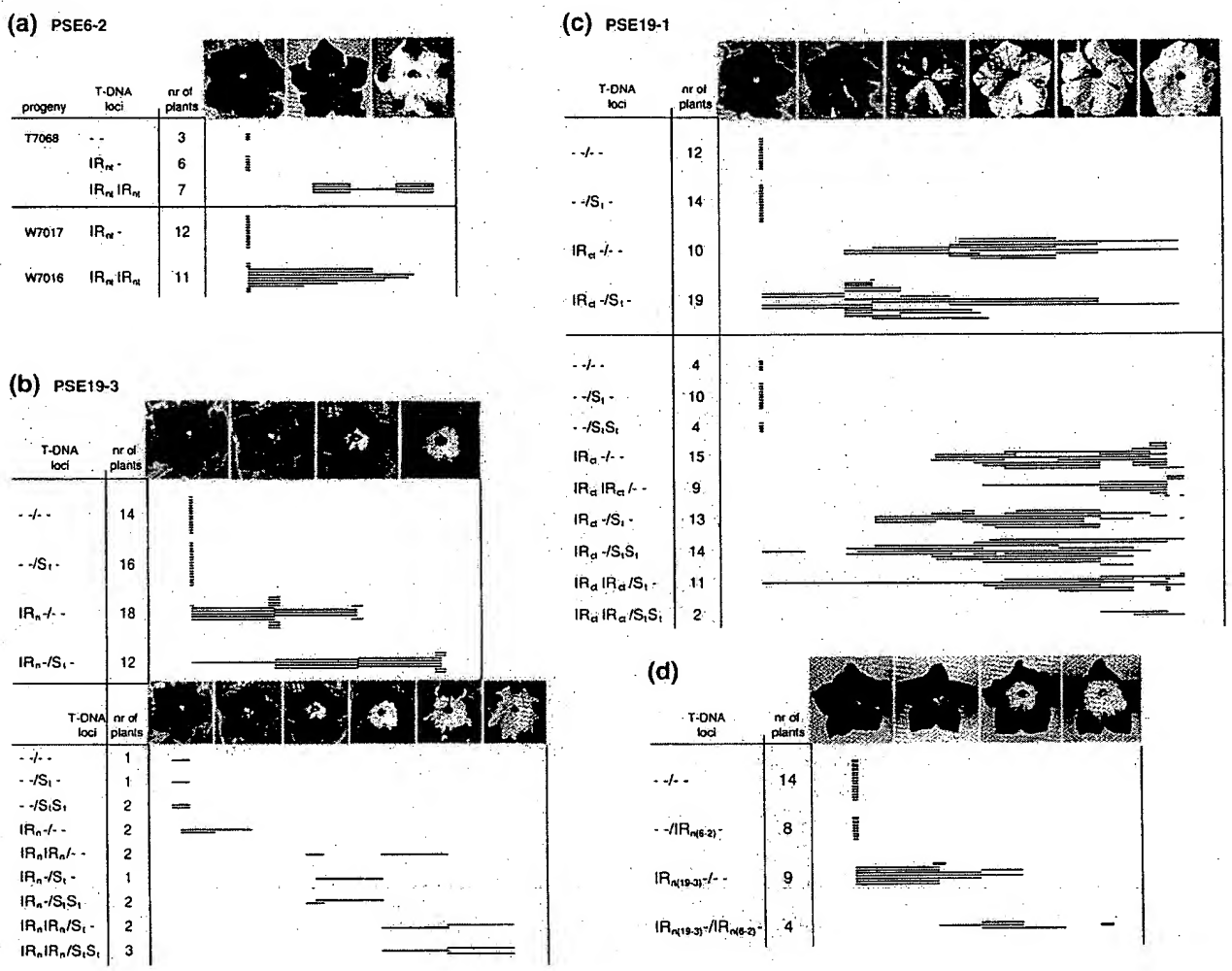


Figure 8. See facing page for legend.

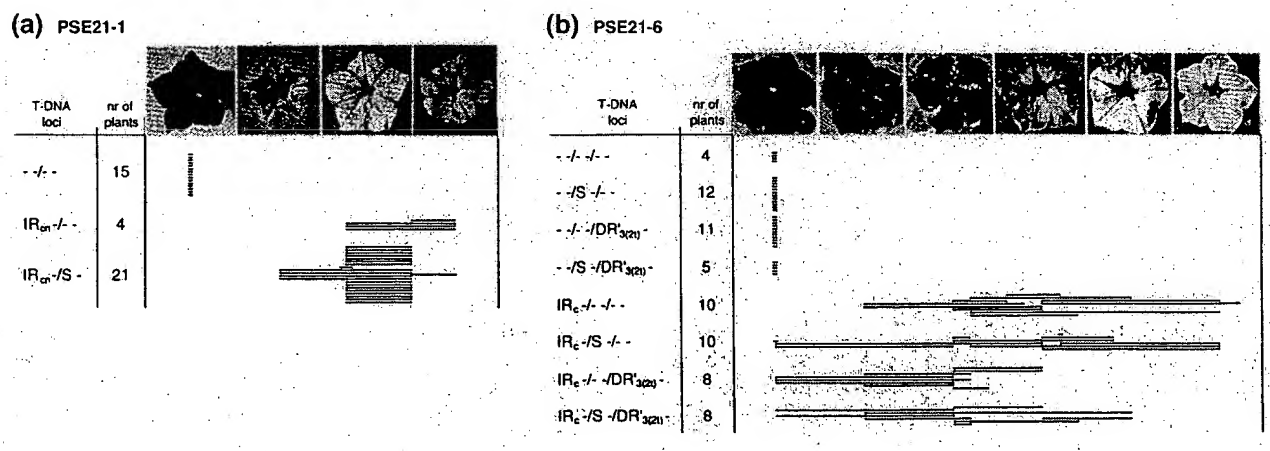


Figure 9. See facing page for legend.

flowers. The S_t locus was not able to induce silencing and its presence seems even to diminish silencing by the IR_{ct} locus. The effect of homozygosity of the IR_{ct} locus was examined in the progeny of a self-fertilization of PSE19-1-4. The progeny (X7030, 82 plants) were examined for T-DNA copy number and phenotype, and as shown in the lower part of Figure 8(c), plants without a T-DNA and plants with just the S_t locus, in a hemizygous or homozygous state, did not show silencing. Again, silencing only occurred in plants containing the IR_{ct} locus and was in homozygotes stronger than in hemizygotes.

Silencing by promoterless ChsA *transgenes is associated with inverted repeats*

We have previously shown that T-DNAs containing a promoterless *ChsA* cDNA can silence *Chs* expression post-transcriptionally (Van Blokland *et al.*, 1994). Analysis of 15 primary transformants indicated that eight contained multiple T-DNAs and that silencing did not occur in any of the plants carrying just monomeric T-DNA insertions (Van Blokland, 1994). The three transformants that produced white or partially white flowers contained multimeric T-DNA loci. We have analysed two of these transformants, PSE21-1 and PSE21-6, in more detail.

PSE21-1. The corollas of PSE21-1 are almost completely white (Van Blokland *et al.*, 1994; Figure 9a). PSE21-1 contains two T-DNA loci: one consists of an intact monomeric T-DNA (S) and the second consists of three intact T-DNAs that are in an inverted orientation relative to one another (IR_{cn}, Figures 7 and 5). The segregation pattern of these two T-DNA loci was examined in a progeny of 75 plants derived from a back-cross (Figure 1b). Twenty-eight plants contained both T-DNA loci, five contained just the IR_{cn} and none contained just the S locus. A χ^2 test indicated that the S and IR_{cn} loci are on the same chromosome, at least 16 cM apart (data not shown). It is not understood why no plants were obtained with just the S locus. Forty of the 75 plants were examined for their silencing phenotype. This revealed that only plants carrying the IR_{cn} locus, alone or together with S, produced white corollas with patches of purple cells, indicating that one IR_{cn} copy is sufficient to confer strong silencing and that the S locus is not necessary (Figure 9a).

PSE21-6. The corollas of PSE21-6 show an erratic distribution of small white sectors (Van Blokland *et al.*, 1994). This transformant carries three T-DNA loci (Figures 7 and 6) which segregate in the progeny of a back-cross in a Mendelian fashion, indicating that the loci are located on different chromosomes. One locus consists of a single intact T-DNA (S) whereas the second locus consists of a two-copy IR locus of which the *Chs* sequences are near the centre (IR_c). The third locus consists of two truncated and one intact T-DNA ordered in a tandem array ($DR'_{3(2t)}$). The truncated T-DNAs are at the borders: one is missing the *NptII* coding region and the other the *ChsA* cDNA. The intact T-DNA and the truncated T-DNAs are separated by complete copies of the pBin19 binary vector.

The role of each of these three T-DNA loci in silencing was examined in a population of 68 plants (W7003) which were derived from a back-cross of PSE21-6 to V26. This revealed that neither the S nor the $DR'_{3(2t)}$ locus, alone or together, were able to confer silencing (Figure 9b). Silencing was only observed in plants carrying the IR_c locus. The S and $DR'_{3(2t)}$ loci appeared to suppress the IR_c-induced silencing. Figure 10 gives a summary of the T-DNA loci conferring silencing.

Suppression and enhancement of IR-induced silencing

Although none of the monomeric T-DNA integrations or the unusual $DR'_{3(2t)}$ locus of PSE21-6 was able to induce

Figure 8. Inheritance of *Chs* silencing with the P_{CaMV}*-UidA-ChsA* containing T-DNA loci.
(a) PSE6-2. The T-DNA locus of PSE6-2 is an IR_{nt} (Figures 7 and 2). Four series of progeny plants, T7001, T7068, W7017 and W7016, obtained as indicated by the crossing scheme in Figure 1b, were examined for T-DNA inserts by Southern blots. The degree of *Chs* silencing was monitored for all the flower corollas produced in the first weeks of flowering. The results of the latter three are presented in the matrix. The names of the different progeny and the genotypes with respect to the T-DNA loci are listed at the left and the number of plants examined are indicated. The top shows the corresponding flower phenotypes. Each bar represents a single plant. Its length indicates the variation in the degree of *Chs* silencing between different flowers from that plant. The parent transformant of the progeny W7016/17 produced white flowers with purple edges and was homozygous for the IR_{nt}.
(b) PSE19-3. S5055-5, which contains the IR_n and S_t (Figures 7 and 3) was crossed with V26 (upper panel) and self-fertilized (lower panel) and the progeny examined as described in (a).
(c) PSE19-1-4 contained an IR_{ct} and a monomeric S_t locus (Figures 7 and 4). The progeny of a back-cross (upper panel) and of a self-fertilization of PSE19-1-4 (lower panel) was examined as above.
(d) The effect of two non-allelic IR loci on *Chs* silencing. A PSE6-2 descendant (T7068-5; IR_{nt}/–) was crossed with a PSE19-3 descendant (S5055-2; IR_n/–) and the progeny examined as described above. Two series of plants were raised: the thick lines represent plants sown four months earlier than those represented by the thin lines. Key: –, no T-DNA; IR– or S–, plants hemizygous for the corresponding T-DNA locus; IR IR or S S, plants homozygous for the relevant T-DNA locus.

Figure 9. Inheritance of *Chs* silencing with the T-DNAs containing the promoterless *ChsA* cDNA.
(a) PSE21-1. PSE21-1 contained an IR_{cn} and an S locus (Figures 7 and 5). Three families, W7008, W7002, and W7036 (Figure 1b) were examined for the T-DNA loci and flower pigmentation phenotypes as described in Figure 8.
(b) PSE21-6. This transformant contained three T-DNA loci, an IR_c, an S and the $DR'_{3(2t)}$ locus (Figures 7 and 6). The primary transformant was back-crossed to V26 and the progeny examined for the T-DNA loci they contained and the pigmentation phenotypes, as described in Figure 8.

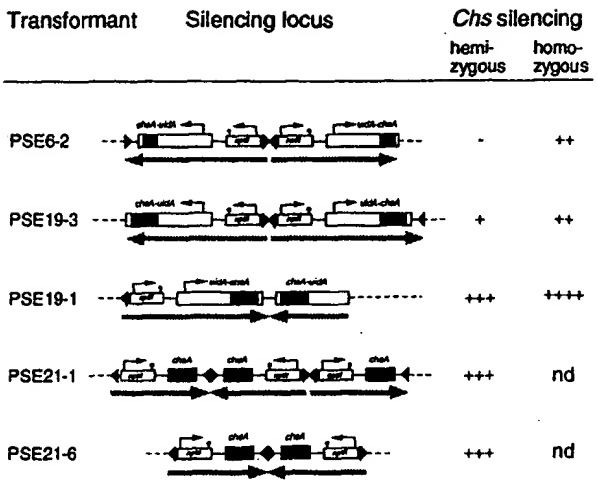


Figure 10. Summary of the type of T-DNA loci able to silence *Chs* expression and the effects of hemi- or homozygosity for the T-DNA loci. nd, not determined. The arrows below the maps denote the palindromic nature and orientation of the integrated transgenes. See text for further details.

silencing, the inheritance studies clearly indicate that they affect the degree of silencing by the IR loci (Figures 8 and 9). The results obtained with this first series of transformants show that silencing by IR_c loci is decreased whereas silencing by an IR_n locus is enhanced by a non-IR locus. The decrease in silencing is observed with an IR_c locus that contains CaMV-35S promoter driven *UidA–ChsA* transgenes (PSE19-1-4, Figure 8c) and with IR_c loci that contain promoterless *ChsA* transgenes (PSE21-1 and PSE21-6, Figures 9a and 9b). This suppressive effect on IR_c-induced silencing is not readily explained but at least indicates that silencing is not activated simply by increasing the number of transgenes. The enhancing effect of the non-IR PSE19-3 locus (S_t) on the PSE19-3 IR_n locus is indicated by an increasing number of plants for which the corollas show a more severe silencing phenotype (Figure 8b). The S_t locus of PSE19-3 retains this enhancing effect on the PSE19-3 IR_n locus after it had been separated from the PSE19-3 IR_n locus for some time and combined again (data not shown). The PSE19-3 S_t locus was, however, unable to activate or enhance the silencing capacity of the IR_n locus of PSE6-2 in IR_n-hemizygous plants, even in plants homozygous for S_t ($IR_{n(6\text{-}2)}$–/$S_tS_{t(19\text{-}3)}$; data not shown). This suggests that the enhancing effect depends in part on features of the IR locus itself and emphasizes that the enhancing effect of the S_t on the IR_n of PSE19-3 is not simply the result of an increase in transgene dosage.

Silencing induced by the PSE19-3 IR_n locus is also enhanced by the IR_{nt} locus of PSE6-2. These two IR loci were combined by crossing the transformant T7068-5 (IR_{nt}/–) with S5055-2 (IR_n/–). Thirty-five progeny plants were analysed for their T-DNA locus and pigmentation phenotype. Figure 8(d) shows that, as observed before (Figure 8a and b), one copy of the PSE6-2 IR_{nt} locus does not induce silencing and one copy of the PSE19-3 IR_n provokes a moderate silencing. However, silencing in plants containing both IR loci is more severe. On average, the white area of the corollas grown on these plants is larger. Silencing by the PSE19-3 IR_n appears dominant as the corolla pigmentation phenotype of these double IR_n transformants is similar to that of the PSE19-3 corollas which have a white ring around the tube rather than having white edges on the limbs as in the PSE6-2 corollas (Figure 8a).

Analysis of a Chs *silencing revertant: loss of the IR_n locus from epidermal L1 cells*

One of the plants that was derived from a back-cross of PSE19-3 to V26, S5055-14 (Figure 3), contained a side branch that produced wild-type pigmented flowers (Figure 11a), indicating that silencing of the *Chs* genes was lost. This revertant branch, termed S5055-14R, was propagated via cuttings and displayed a stable wild-type phenotype. In the original transformant, the IR_n locus was shown to be responsible for the silencing of *Chs* (Figure 8b). This raised the question whether the IR_n locus was lost or rearranged in the revertant, which would explain the loss of silencing. To test this possibility, we analysed DNA from corollas of the parental and revertant plant by digesting the DNA with *Hin*dIII and hybridizing the Southern blots with a *UidA* probe which detects the S_t and IR_n-specific fragments (Figure 11b). This showed that corollas from the parental plant contained the known S_t and IR_n-specific fragments and as expected the bands were of equal intensity (lane 1). In contrast, the corollas of the revertant did contain the IR_n-specific bands but their intensity was much lower than those of the parent (lane 4), while the S band intensity was the same as that of the parent. These results indicated that the IR_n locus was indeed affected in the revertant. One of the possibilities was that the IR_n locus was deleted in a fraction of the corolla cells. We therefore analysed DNA from other tissues which showed that in DNA of leaves (lane 5) and stem (lane 6) from the revertant, the IR_n and S_t-specific bands were of equal intensity, similar to those of the parent (lanes 2 and 3). Given this observation, one would not infer the IR_n locus to be lost in the revertant. However, plant tissues are composed of three meristematic layers, L1, L2 and L3 (Huala and Sussex, 1993), and one of the differences between a corolla, a stem, and a leaf is that the ratios of L1, L2 and L3-derived cells in these tissues are different. As compared with stem and leaf, corollas contain a much larger proportion of L1 epidermal cells. Thus, loss of the IR_n locus from L1 cells could explain the reduced hybridization intensity on Southern blots of corolla DNA compared with stem and leaf DNA. Furthermore, it would also explain the reversion to wild-type flowers as the *Chs* gene is predominantly expressed in the L1 epidermal cells of the corolla where it

is involved in synthesis of the anthocyanins (Martin and Gerats, 1993). This possibility was tested by analysing genomic DNA obtained from L1 cells for which we used trichomes. These trichomes were harvested from stems of the parent and revertant and the DNA was analysed as described above. Figure 11(c) shows that trichomes from the parental plant contain both the IR_n- and S_t-specific bands, which are of equal intensity. However, trichome DNA from the revertant only contained the S_t-specific fragment and no trace of IR_n fragments. This result is consistent with the specific loss of the IR_n locus from the L1 cells of the revertant branch.

As the gametes are of L2 origin, it was possible to verify that the IR_n locus was present in L2 cells by following the segregation of S_t and IR_n in the progeny of a cross between the revertant and untransformed V26. If the L2 cells had also lost the IR_n locus, then it would of course not be transmitted to the progeny. For this, 26 progeny plants (W7050) were analysed for their T-DNA genotype by Southern blotting, and Figure 11(d) shows the results for 15 of these. This Southern blot indicates that the IR_n-specific fragments are transmitted to the progeny and that they segregate in a Mendelian manner which is expected if L2 cells contain the IR_n locus. The flowers of the plants containing the IR_n locus showed a *Chs*-silencing phenotype similar to that of the parent. The reversion is therefore not heritable, as expected for an L1-specific trait. Taken together, the analysis of this somatic reversion shows the importance of the IR_n locus in silencing in this plant. Furthermore, as the IR_n was only deleted from the L1 cells, this result indicates that silencing cannot be induced by neighbouring IR_n-containing L2 cells.

Fluorescence *in situ* hybridizations indicated that the IR_n locus is near the telomere of chromosome 4 (Fransz *et al.*, 1996; unpublished results). The nearby chromosome 4-specific genes *DfrA* and flavanone 3-hydroxylase (*F3h*) used as probes on the blot of Figure 11(b), also gave rise to a lower hybridization intensity of the corresponding gene fragments in the flowers of the revertant (not shown). However, flow cytometry analysis on nuclei isolated from

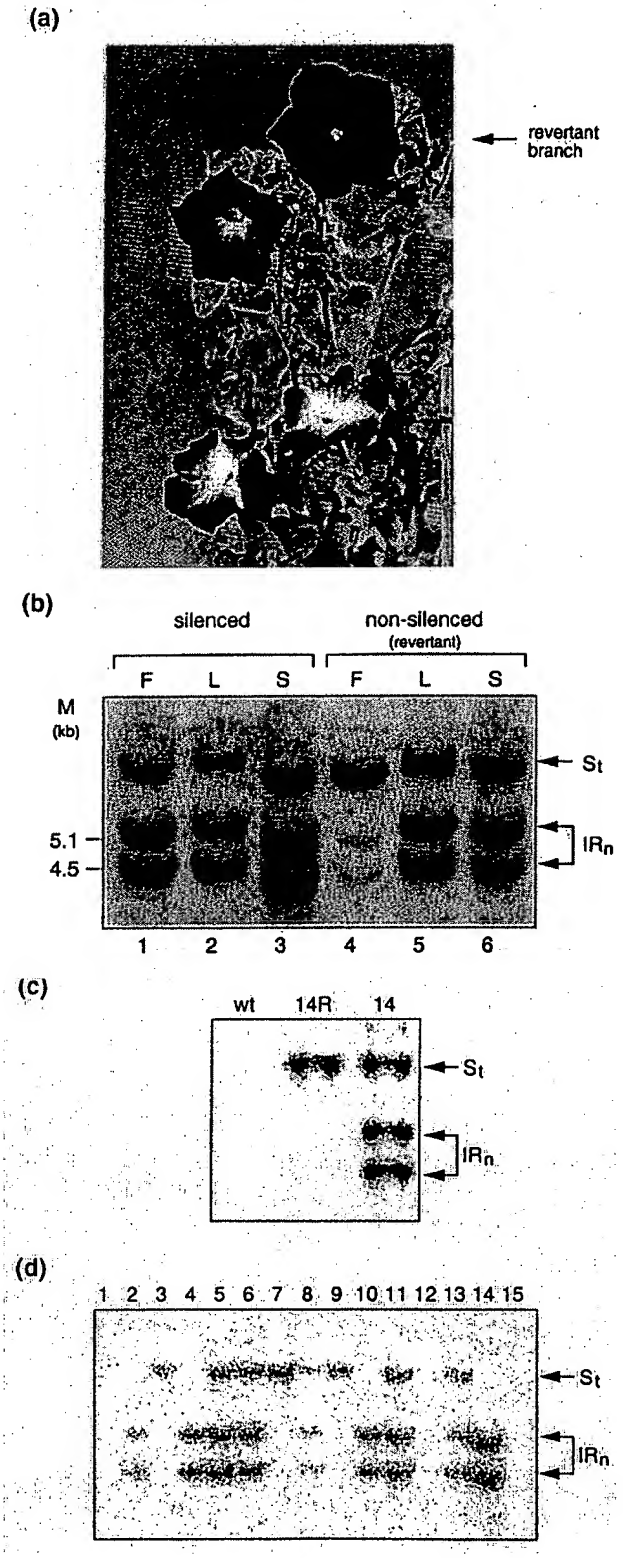


Figure 11. Analysis of *Chs* silencing revertant: specific loss of IR_n locus from L1 cell layer.

(a) The S5055-14 transformant produces flowers in which *Chs* expression is silenced near the tube. After pruning, a side branch emerged that produced normally coloured flowers indicating loss of silencing. This revertant branch was further propagated by cuttings and was named S5055-14R.

(b) Southern blot of *Hin*dIII-digested DNA of corollas (F), leaves (L) and stem (S), from the original transformant (lanes 1–3) and from the silencing revertant (lanes 4–6), which was hybridized with a *UidA* probe. The position of the fragments derived from the single-copy insert (S) and from the IR_n insert (IR) are indicated at the right.

(c) Southern blot of DNAs isolated from stem trichomes which are from L1 origin: from an untransformed plant (wt), silencing revertant (14R) and original transformant (14). The DNA was digested with *Hin*dIII and hybridized with a *UidA* probe.

(d) *UidA* hybridization of a Southern blot that contained *Hin*dIII-digested DNA of progeny plants derived from a cross of the *Chs* silencing revertant S5055-14R to untransformed V26. A sample of 15 out of 26 plants is shown. This revealed that L2 cells of the revertant contain the IR_n locus.

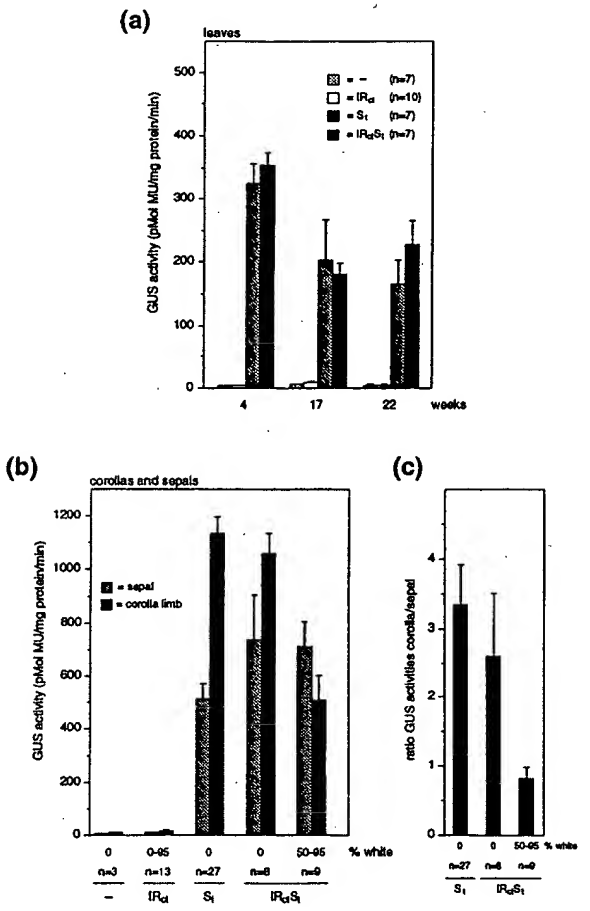


Figure 12. Corolla-specific silencing of a monomeric *UidA–ChsA* transgene by a silent IR_{ct} locus.
The GUS enzyme activities were measured in individual leaves (a) and in individual corolla limbs and sepals (b) from untransformed V26 (–), from transformants hemizygous for the IR_{ct} locus of PSE19-1 (IR_{ct}), from transformants hemizygous for the S_t locus of PSE19-3 (S_t), and from transformants hemizygous for both loci ($IR_{ct}S_t$). The GUS activities in leaves were measured 4, 17 and 22 weeks after sowing. The activities in corollas and sepals were measured 32–34 weeks after sowing. Panel (c) shows the mean of the ratios of the GUS activities in corolla limbs and sepals (C/S ratio) of flowers from S_t and $IR_{ct}S_t$ transformants. The degree of *Chs* silencing in the corollas is indicated by the percentage of the corolla that was white. As it was impossible to separate the pigmented sectors from the white sectors the whole corolla was used to prepare the extract. GUS activities are expressed as mean ± standard error of the mean (SEM); n is the number of samples analysed. Analysis of variance indicated a highly significant difference in the C/S ratios of the purple S_t flowers and the 50–95% white $IR_{ct}S_t$ flowers ($P < 0.0001$).

the revertant did not show a detectable reduction in DNA content, suggesting that just a small part of chromosome 4 was deleted including the IR_n, *DfrA* and *F3h* genes. Whether this deletion is due to the presence of the IR is unknown, but since we have observed this type of somatic reversion only once, it appears that the IR_n locus is quite stable, in contrast to long inverted repeats in mammals (Collick *et al.*, 1996). Small rearrangements at the junction between the T-DNAs, rendering IR structures more stable (Collick *et al.*, 1996; Leach, 1994), cannot be excluded.

Silencing by the PSE19-1 IR_c locus requires an active endogenous Chs *gene*

Inverted repeat loci with *Chs* sequences near the IR centre suppress the expression of endogenous *Chs* genes relatively strongly. As the IR_c *Chs* sequences are not transcribed or only at a very low level (Van Blokland *et al.*, 1994), silencing by these loci cannot be explained by the RNA threshold model. However, it is possible that in most plant tissues these IR_c loci produce low quantities of aberrant transcripts sufficient to activate the RNA degradation machinery (English *et al.*, 1996). If this is the case, it is expected that, for example, the IR_{ct} locus of PSE19-1, for which the *UidA–ChsA* transgenes are not detectably expressed (Figure 12a and b), will silence the monomeric *UidA–ChsA* transgene of the PSE19-3 S_t locus (Figure 7) in other tissues than the corolla. The *UidA–ChsA* transgene of S_t is expressed in leaves giving rise to a clearly detectable GUS activity (Figure 12a and b). An alternative possibility is that the IR_{ct} locus requires an active endogenous *Chs* gene in order to induce silencing. This would mean that suppression of the PSE19-3 S_t *UidA–ChsA* transgene only occurs in cells where the *Chs* genes are highly transcribed, such as the epidermis of the corolla. To distinguish between these alternatives, expression of the PSE19-3 S_t *UidA–ChsA* transgene was determined in the presence or absence of the PSE19-1 IR_{ct} locus in leaves, sepals and corollas. These plants were obtained by crossing an $S_{t(19\text{-}3)}$ plant (S5055-8) with an $IR_{ct(19\text{-}1)}$ plant (W7001-58).

In the absence of the IR_{ct} locus, the S_t gene is clearly expressed in leaves (Figure 12a, S_t, hatched columns) and in sepals (Figure 12b, S_t, hatched columns). Expression in these tissues is not reduced by the IR_{ct} locus (Figures 12a and b, $IR_{ct}S_t$). Even in leaves of older plants (22 weeks), this expression is not detectably influenced by the IR_{ct} locus (Figure 12a). Thus, although the IR_{ct} *UidA–ChsA* transgenes are not expressed or only at very low levels (Figure 12a, white columns), due to transcriptional silencing (not shown), this locus is unable to silence the homologous S_t *UidA–ChsA* transgene, either transcriptionally or post-transcriptionally. This indicates that there are no aberrant transcripts derived from the PSE19-1 IR_{ct} locus that can induce the PTGS mechanism. The results obtained with corollas, in which *Chs* is silenced by the IR_{ct} locus, are different. In the absence of the IR_{ct} locus, the S_t *UidA–ChsA* gene in corollas is expressed about threefold higher than in sepals (Figure 12c) and leaves (compare Figure 12a and b). In addition to *Chs*-silenced flowers, the $IR_{ct}S_t$ plants also produce fully purple corollas, indicating that *Chs* expression was not suppressed. In these purple corollas, the S_t *UidA–ChsA* gene is expressed about threefold higher than in sepals (Figure 12c, $IR_{ct}S_t$, 0% white) and leaves (compare Figure 12a and b). However, in corollas that contain large white sectors with randomly distributed

purple cells, the expression is fourfold reduced (Figure 12c, $IR_{ct}S_t$, 50–95% white) as compared with that in purple corollas with just the S_t locus. Taken together, these results show that down-regulation of the S_t transgene by the IR_{ct} locus only occurs in tissues in which the endogenous *Chs* genes are normally highly active, and moreover, only when they are post-transcriptionally silenced. This co-ordinate silencing of the S_t locus and the endogenous genes by the IR_{ct} locus thus appears to require transcription of the endogenous *Chs* genes. That the S_t *UidA–ChsA* gene is not completely silenced can be attributed to the fact that it might still be expressed in the L2 cell layer which is sandwiched between the upper and lower epidermis and which is not transcribing *Chs*, and the fact that the flowers tested still contained 5–50% purple cells in which the *UidA–ChsA* transgene of the S_t is probably expressed as it is in purple flowers (Figure 12c). It was not possible to verify this because the GUS activity was too low to detect it histochemically. These results are consistent with a model in which aberrant RNAs, thought to be necessary to induce the RNA degradation machinery, are derived from one or more of the endogenous genes. We infer that this happens as a consequence of an ectopic interaction with the IR locus. Why the IR_{ct} locus does not inactivate the S_t locus in this way is not understood.

Discussion

To obtain an insight into the mechanism(s) of post-transcriptional silencing of endogenous genes, we have identified and physically characterized the T-DNA loci responsible for the silencing of endogenous *Chs* genes in a series of previously described transformants (Van Blokland *et al.*, 1994). Our results show that silencing is associated with the presence of multimeric T-DNA loci in which the T-DNAs, which harbour the transgenes, are arranged as inverted repeats. This was found for the chimaeric *UidA–ChsA* transgenes driven by the CaMV 35S promoter and for the promoterless *ChsA* transgene.

Structure of the T-DNA integrations

In the transformants showing gene silencing, various types of T-DNA integrations were found (Figure 7), but all contained a T-DNA locus that was composed of two or more T-DNAs arranged as inverted repeats. PSE6-2 and PSE19-3 contained an IR_n type locus with the *NptII* genes near the centre of the IR and *Chs* sequences distal to the centre, whereas PSE19-1 and PSE21-6 contained an IR_c T-DNA locus with the *Chs* sequences proximal to the IR centre. PSE21-1 contained a locus composed of three T-DNAs that are arranged as inverted repeats (IR_{cn}, Figure 7). Two T-DNA loci, $DR'_{3(2t)}$ of PSE21-6 and S_t of PSE19-3 (Figure 7), contained DNA from the binary vector pBin19 which was used for the transfection. Co-transfection of vector DNA appears to occur rather frequently (Martineau *et al.*, 1994). However, as these vector DNA-containing T-DNA loci did not segregate with the silencing phenotype, these sequences are not involved in silencing. Many of the T-DNAs of both the monomeric and multimeric loci are truncated, either at the left border or at the right border. The breakpoints of these partial T-DNAs have not been precisely mapped by sequencing. However, the Southern blot analyses and the use of the various probes provided sufficient information about the parts that are missing. As silencing was not associated with the presence of a particular type of truncated T-DNA, it is unlikely that partial T-DNAs play a role in establishing silencing.

Silencing of Chs *requires the presence of an IR locus*

Silencing of *Chs* expression coincides with the presence of an IR locus (Figures 8 and 9), indicating that such a locus is important for activating the process. This is supported by the results obtained with the somatic revertant which shows that a deletion of the IR locus from L1 cells results in loss of silencing in these cells (Figure 11). Furthermore, a survey of our entire collection of transgenics so far indicates that in addition to the characterized transformants analysed in this study, 26 other transformants that contained silenced endogenous genes contain an IR locus or a more complex locus. In contrast, none of the monomeric T-DNA copies or the $DR'_{3(2t)}$ locus of PSE21-6 conferred silencing (Figure 8 and 9) and 43 other transformants containing one or more monomeric T-DNA integrations also do not show silencing of the endogenous genes (unpublished results). This compilation and the segregation data presented in Figures 8 and 9 indicate that the structural organization of a transgene locus is important for activating the PTGS mechanism. In some other studies, PTGS was also found associated with multimeric transgene loci (De Carvalho Niebel *et al.*, 1995; Depicker *et al.*, 1996; English *et al.*, 1996; Hobbs *et al.*, 1993; Kunz *et al.*, 1996). However, the exact structural organization of the locus was not determined in all cases, and also the importance of the repetitive character of the silencing loci was not emphasized. In the case of *NptII* (Depicker *et al.*, 1996) and *UidA* silencing (English *et al.*, 1996; Hobbs *et al.*, 1993), the T-DNAs were in an IR configuration. Jorgensen *et al.* (1996) also observed silencing of *Chs* in petunia by IR loci.

However, several studies show PTGS associated with a single monomeric T-DNA locus (Dorlhac de Borne *et al.*, 1994; Elmayan and Vaucheret, 1996; Jorgensen *et al.*, 1996, Palauqui and Vaucheret, 1995) which raises the question about the relevance for multimeric T-DNA loci in activating the PTGS mechanism. In two of these cases, the transgenes were expressed from an enhanced CaMV 35S promoter (Elmayan and Vaucheret, 1996; Jorgensen *et al.*, 1996)

which seems to cause suppression in all or most of the transformants. In the other two cases, a regular 35S promoter was used and the frequency with which silencing was observed was much lower than with the enhanced promoter. These results indicate that the higher the amount of transgene RNA accumulation, the higher the silencing frequency, which is consistent with the RNA-threshold model of PTGS. The *UidA–ChsA* transgenes in our constructs were also transcribed from the regular 35S promoter but we have not observed silencing by monomeric T-DNAs. One difference is that the transcripts from our chimaeric *UidA–ChsA* genes hardly accumulate despite the fact that the transgenes are sometimes highly transcribed, as determined by run-on assays (Van Blokland *et al.*, 1994). Apparently these transcripts are intrinsically unstable and therefore may not reach the proposed threshold level.

A particular threshold level can also be reached by a high transgene copy number or by increasing the number of transgenes by crossings, by combining ectopic loci or by making plants homozygous (Angenent *et al.*, 1993; De Carvalho *et al.*, 1992; De Carvalho Niebel *et al.*, 1995; Dorlhac de Borne *et al.*, 1994; Hart *et al.*, 1992; Palauqui and Vaucheret, 1995; Vaucheret *et al.*, 1995). We also observed such gene dosage effects. Plants homozygous for silencing loci and plants containing two non-allelic silencing loci show a more severe silencing phenotype than plants carrying a single silencing locus (Figure 8). These findings can be interpreted in two ways. The first is that a higher gene dosage results in a higher production of transgene RNA which is responsible for triggering the PTGS mechanism via the RNA threshold mechanism. Indeed, the silencing sequences of the IR_n loci are transcribed (Van Blokland *et al.*, 1994). However, monomeric *UidA–ChsA* transgenes in a homozygous state can be transcribed as high as those of a single IR_n locus and yet do not induce silencing. Thus the amount of transcripts *per se* does not seem important. This is supported by the fact that strong IR_c loci are barely transcribed, if at all. Another relevant observation is that the IR_{ct} locus of PSE19-1 does not silence by itself but appears to require the endogenous *Chs* gene (Figure 12). We therefore favour a second possibility, in which the palindromic arrangement of the silencing sequences within the IR loci plays a crucial role (see also below). How the effects of monomeric T-DNA loci on the IR loci fit in is not understood. With one locus, we observed enhancement of IR_n-induced silencing (Figure 8), whereas with the others we saw a reduction in silencing by IR_c loci. Whatever the underlying mechanisms of these opposite effects, these findings are not easy explained by current RNA threshold or gene dosage models.

To correlate the seemingly contradictory results with the IR loci described here and the monomeric loci described by others (Dorlhac de Borne *et al.*, 1994; Elmayan and Vaucheret, 1996; Jorgensen *et al.*, 1996, Palauqui and Vaucheret, 1995), information is required about the fate of the transcripts from the endogenous genes and/or the transgenes. It has been proposed that some kind of aberrant RNA activates or catalyses the degradation of specific transcripts (English *et al.*, 1996; Smith *et al.*, 1994). Following this line of reasoning, it is conceivable that there are different ways by which such an RNA species might be produced: (i) via the excessive production of stable RNA, by using a strong promoter driving the transgenes (Elmayan and Vaucheret, 1996; Goodwin *et al.*, 1996; Jorgensen *et al.*, 1996; Metzlaff *et al.*, 1996; Smith *et al.*, 1994); (ii) by the expression of transgenes that are modified (Ingelbrecht *et al.*, 1994) and/or located in repeats (Depicker *et al.*, 1996); and (iii) by the endogenous gene(s) when their expression is altered by means of a (transient) ectopic interaction with the IR locus (see below). Such an interaction may only be possible if the transgene locus is repetitive, and perhaps more important, the silencing sequences close to the centre of an IR. If structural properties of a silencing transgene locus are indeed the most important features, it is evident that the transgenes may not have to be highly transcribed, if at all (Van Blokland *et al.*, 1994), which would explain the efficient silencing by IR_c loci carrying promoterless *Chs* sequences.

Differences between IR_c or IR_n type loci

Although IR_c and IR_n type loci both induce silencing, they display some differences. Firstly, silencing by an IR_c is more severe than by an IR_n (Figures 8, 9 and 10). Secondly, the silencing capacity of an IR_n locus declines in successive generations (Figures 8a; unpublished results), while that of IR_c loci appears more stable. Finally, the distribution of silenced (white) cells in the corolla seems different in IR_n- and IR_c-containing corollas (Figures 8 and 9). The white sectors of IR_n corollas have a fairly regular pattern whereas those of IR_c corollas are more erratic. Jorgensen *et al.* (1996) also reported differences in pigmentation patterns in petunia flowers that were correlated with differences in the repetitiveness and organization of the transgene loci. Hardly anything is known about the formation of these patterns but it seems unlikely that local differences in the transcriptional activity of the endogenous genes are responsible (Jorgensen, 1995). If this is true, every silencing transformant is expected to have the same basic pigmentation phenotype which is clearly not the case. The type of transgene locus seems to determine the type of variegated pigmentation pattern. Taken together, these results suggest that IR_n and IR_c loci may activate silencing along different pathways, which appears to be related to the different positions of the silencing sequences within the IR.

IR loci and aberrant transcripts

In several cases of post-transcriptional silencing, the level of transgene expression is also not directly correlated with the degree of silencing (English *et al.*, 1996; Kunz *et al.*, 1996; Mueller *et al.*, 1995). It has therefore been proposed that for activating the RNA degradation activity, a fraction of the transgene transcripts has to be aberrant (Baulcombe and English, 1996; Dougherty and Parks, 1995) in structure, base modification, or the degree of processing (Metzlaff *et al.*, 1996; Van Blokland *et al.*, 1996). The plant-encoded RNA-dependent RNA polymerase (RdRP; Lindbo *et al.*, 1993) may use these aberrant or excessively produced RNAs as a template and would synthesize complementary RNAs (cRNA, or antisense RNA; Dougherty and Parks, 1995) which in turn would tag other complementary RNAs for degradation by dsRNA-specific ribonucleases. A role for cRNAs, produced by the RdRP, is attractive as it explains the sequence specificity of the PTGS mechanism.

Do the characterized IR silencing loci produce such aberrant transcripts? It is unlikely that the IR_c loci carrying the promoterless *Chs* transgenes produce aberrant transcripts as these sequences are not detectably transcribed. Moreover, there is no detectable read-through transcription from one repeat into the other. The T-DNA of the pBin19 vector used to generate transgenic petunia plants contains M13 DNA (Fray *et al.*, 1994). Hybridization of labelled nascent RNA obtained by nuclear run-on transcription to M13 vector DNA did not result in signals above background levels (Van Blokland *et al.*, 1994). It is furthermore unlikely that the detected antisense transcripts from the CaMV 35S promoter-driven *UidA–ChsA* transgenes (Van Blokland *et al.*, 1994) provoke silencing because the levels are so low, and monomeric T-DNAs can produce as much antisense RNA or even more and yet not induce silencing (data not shown). As a result of specific characteristics of an IR locus, transcription of genes within such a locus could potentially result in aberrant RNAs. However, if a low level of IR-derived aberrant RNAs were responsible for activating the silent state, then silencing of the $S_{t(19\text{-}3)}$ *UidA–ChsA* transgene would be expected in leaves, for example. This is not observed (Figure 12a). In partially white corollas, however, where the endogenous genes are transcriptionally highly active but post-transcriptionally silenced (Figures 12b and c), the expression of the $S_{t(19\text{-}3)}$ *UidA–ChsA* transgene is fourfold lower, indicating that the gene is down-regulated. This suggests that the endogenous genes play a key role in the post-transcriptional silencing process.

One possibility is that the endogenous genes produce the aberrant RNA species. An elevated level of unspliced *Chs* transcripts in nuclei containing post-transcriptionally silenced *Chs* genes suggests that the normal production of *Chs* mRNA is to some extent impaired (Van Blokland *et al.*, 1996). It is therefore tempting to speculate that an IR locus, and in particular the IR_c locus, is able to interact at some point during corolla development with the endogenous *Chs* gene(s) via DNA–DNA pairing (Baulcombe and English, 1996; Jorgensen, 1992; Van Blokland *et al.*, 1994), thereby interfering with the normal processing and/or transport of transcripts and hence generating possibly aberrant RNAs.

Possible role of IR loci in PTGS: DNA pairing

Inverted repeats are known to be a source of genomic instability in prokaryotes (Bi and Liu, 1996; Leach, 1994) and in eukaryotes (Collick *et al.*, 1996; Gordenin *et al.*, 1993; Henderson and Petes, 1993; Ruskin and Fink, 1993). In contrast in plants, IR loci composed of two or three tandemly inverted repeats, each repeat 4.5 kb or more in length, appear stable. Except for one special case in which the IR_n locus and part of the chromosome was specifically lost from L1 cells (Figure 11), which seems unrelated to the IR locus itself, we have no indications for gross DNA rearrangements. Small rearrangements at the centre of the IRs, which can lead to a more stable IR (Collick *et al.*, 1996; Leach, 1994) cannot be excluded.

In *Drosophila*, closely linked repeats, including inverted repeats, of a P transposon carrying a *white* transgene tend to become silenced by means of heterochromatin formation and which gives rise to *white* variegation (Dorer and Henikoff, 1994). It was proposed that pairing of the closely linked repeats may result in the formation of folded structures that are recognized by heterochromatic proteins. By analogy, similar interactions may occur between the sequences within the plant IRs, and although the *Chs* transgenes at the boundaries of the IR are still active, it is striking that the *Chs* genes near the centre are mostly inactive (Van Blokland *et al.*, 1994). This inactivation is associated with an increased methylation (unpublished results) but whether these genes have a condensed chromatin structure is as yet unknown. IR structures might be prone to pair with one or more of the ectopic homologous endogenous genes, which may occur even without strand displacement (Camarini-Otero and Hsieh, 1993). In this context, it is interesting to note that, in yeast, IRs create hot spots for mitotic interchromosomal recombination with single-copy sequences (Gordenin *et al.*, 1993) indicating that palindromic DNA senses homologous sequences more easily than non-palindromic DNA, which might be related to the potential stem structures of IRs (Gordenin *et al.*, 1993). Evidence in plants that homologous sequences sense each other and possibly pair, comes from studies of transgene loci of which the pattern of methylation is transferred to unlinked homologous transgenes (Ingelbrecht *et al.*, 1994; Matzke *et al.*, 1994; Matzke *et al.*, 1989; Meyer *et al.*, 1993; Vaucheret, 1993). How this

happens is as yet unknown but may involve a gene conversion-like mechanism. However, for the silenced endogenous *Chs* genes, we have no evidence for rearrangements or changes in methylation.

The presumed pairing between IR sequences and endogenous gene(s) might be stimulated by a particular chromatin structure of the IRs. It is attractive to propose a role for chromatin because a transient interaction early in corolla development may mark or imprint the endogenous gene which later during development may alter its expression and lead to the production of aberrant transcripts. The differential silencing capacities of IR_n and IR_c loci (Figures 8 and 9) might be explained by assuming that the chromatin structure near the centre of an IR is different from that at the borders. Although the effect of the non-silencing locus on an IR_n locus is not readily explained, it is conceivable that in the case of S loci that reduce silencing by an IR_c locus, an increasing number of homologous sites in the genome may cause some kind of competition with the IR.

Future experiments are required to obtain a better understanding of the special features of IR loci and to obtain direct evidence for the proposed DNA–DNA pairing as was elegantly shown for the $Brown^{dominant}$ allele in *Drosophila* (Csink and Henikoff, 1996; Dernburg *et al.*, 1996).

Experimental procedures

T-DNA constructs and plant material

The *ChsA* T-DNA constructs pSE19, pSE6 and pSE21 and the corresponding petunia V26 transformants have been described by Van Blokland *et al.* (1994). The physical maps of the T-DNA constructs are shown in Figure 1.

DNA manipulations

DNA for the Southern blot analyses was extracted from leaves, stem, corollas or trichomes (Dellaporte *et al.*, 1983) and further purified by N-cetyl-N,N,N-trimethyl ammonium bromide (CTAB) precipitation. The trichomes were harvested by putting stems into liquid N_2 after which the frozen trichomes could be removed by a razor blade. DNA for the PCR analysis was obtained from seedlings (Klimyuk *et al.*, 1993). For the Southern blots, 5–10 μg of genomic DNA was digested overnight with the appropriate enzymes, separated on a 0.8% agarose gel at low voltage and transferred onto a Hybond-N^+ membrane (Amersham) by capillary blotting, followed by alkali fixation. The filters were hybridized at 60°C for about 20 h in 10% dextran sulphate, 1% SDS, 50 mM Tris pH 7.5, 1 M NaCl; 0.1 mg sheared herring sperm DNA/ml, containing a double stranded ^{32}P-labelled DNA probe. After the hybridizations, the filters were washed in 0.18 M NaCl, 10 mM NaPi, 1 mM EDTA, pH 7 (SSPE) buffer with a final wash in 0.1 SSPE 0.1% SDS at 65°C for 5 min. The hybridizing fragments were visualized by autoradiography or by using a Phosphor-Imager. Before re-hybridizations, the filters were washed in 0.5% SDS at 100°C for 5 min.

The *UidA* and *ChsA* (+79 to 1413) probes were *Bam*HI fragments, the CaMV 35S promoter probe was a 850 bp *Hin*dIII–*Bam*HI fragment and the nos poly adenylation region probe was a 253 bp *Bam*HI–*Eco*RI fragment, all derived from construct pSE19 (Van Blokland *et al.*, 1994) The pBin19 vector probe was a mixture of two *Eco*RV fragments (2736 bp and 1801 bp) and three *Dra*I fragments (1177 bp, 548 bp and 2932 bp), which together cover the entire pBin19 vector region (Frisch *et al.*, 1995). The ploidy level of the T-DNA loci in progeny from self-fertilizations was determined by Southern blot analysis using the *Chi* or *Fls* probes as internal controls. The band intensities of the T-DNA fragments were compared with those of the *Chi* and *Fls* bands. The reliability of this method was tested by analysing the progeny of a back-cross of a few plants to V26 using PCR of seedling extracts (data not shown). The primers we used were: RB2 (5′-GGAAGCTTTGCT-GGTGGCACGG-3′) and ME1 (5′-GGGATCCGTTGTACGTGCTCTTA-TTGG-3′) which are directed against the nucleotides +1831 to +1852 and +2802 to + 2783 relative to the first nucleotide of the ATG of the *ChsA* gene, and RBO (5′-CGCAAGACCGGCAACAGG-3′), which is directed against the transgenic nos polyadenylation region. The primer combination RB2/RBO amplified a transgene fragment, while the primer combination RB2/ME1 amplified a 972 bp fragment from the endogenous *ChsA* gene, which served as an internal control.

Fluorometric GUS assay and statistical analysis

GUS enzyme activities in the extracts of young leaves, young flower limbs and the corresponding sepals were determined by the fluorometric assay as described by Jefferson *et al.* (1987). The tissues were ground in liquid N_2 in the presence of Dowex-1 (Sigma) to remove flavonoids (Van Tunen *et al.*, 1990). For each flower, the GUS activity of the limb, which was normalized to the protein concentration, was divided by that of the sepal (C/S ratio). There were no differences in protein content between purple and white corollas. To test for significant differences of the C/S ratios between groups, analysis of variance was used on logarithmic values of the C/S ratios, followed by an *a posteriori* comparison with Bonferoni correction.

Acknowledgements

We thank Ada Viterbo for contributing to some of the experiments, and Pieter Hoogeveen and Martina Meesters for taking care of the transgenic plants and for the numerous crosses. We are grateful to Ronald Koes for suggestions on analysing the *Chs* silencing revertant and to Jacques Bedaux for the statistical analysis of the GUS activities. This work was supported by the Netherlands' Technology Foundation (STW), by a research grant to M.S. from S&G Seeds BV, Enkhuizen, The Netherlands, and by a research fellowship from the Royal Netherlands Academy of Sciences (KNAW) to J.M.K.

References

Angenent, G.C., Franken, J., Busscher, M., Colombo, L. and Van Tunen, A.J. (1993) Petal and stamen formation in Petunia is regulated by the homeotic gene-fbp1. *Plant J.* **4**, 101–112.

Baulcombe, D.C. and English, J.J. (1996) Ectopic pairing of homologous DNA and post-transcriptional silencing in transgenic plants. *Curr. Opin. Biotechnol.* **7**, 173–180.

Bi, X. and Liu, L.F. (1996) DNA rearrangement mediated by inverted repeats. *Proc. Natl Acad. Sci. USA*, **93**, 819–823.

Camarini-Otero, R.D. and Hsieh, P. (1993) Parallel DNA triplexes,

homologous recombination, and other homology-dependent DNA interactions. *Cell*, **73**, 217–223.

Collick, A., Drew, J., Penberth, J., Bois, P., Luckett, J., Scaerou, F., Jeffreys, A. and Reik, W. (1996) Instability of long inverted repeats within mouse transgenes. *EMBO J.* **15**, 1163–1171.

Csink, A.K. and Henikoff, S. (1996) Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature*, **381**, 529–531.

De Carvalho, F., Gheysen, G., Kushnir, S., Van Montagu, M., Inzé, D. and Castresana, C. (1992) Suppression of β-1,3-glucanase transgene expression in homozygous plants. *EMBO J.* **11**, 2595–2602.

De Carvalho Niebel, F., Frendo, P., Van Montagu, M. and Cornelissen, M. (1995) Post-transcriptional cosuppression of β-1,3-glucanase genes does not affect accumulation of transgene nuclear mRNA. *Plant Cell*, **7**, 347–358.

De Haan, P., Gielen, J.J.L., Prins, M., Wijkamp, I.G., Van Schepen, A., Peters, D., Van Grinsven, M.Q.J.M. and Goldbach, R. (1992) Characterization of RNA-mediated resistance to tomato spotted wilt virus in transgenic tobacco plants. *Biotechnology*, **10**, 1133–1137.

Dehio, C. and Schell, J. (1994) Identification of plant genetic loci involved in a posttranscriptional mechanism for meiotically reversible transgene silencing. *Proc. Natl Acad. Sci. USA*, **91**, 5538–5542.

Dellaporte, S.L., Wood, J. and Hicks, J.B. (1983) A plant DNA minipreparation: version II. *Plant Mol. Biol. Rep.* **1**, 19–21.

Depicker, A., Ingelbrecht, I., Van Houdt, H., De Loose, M. and Van Montagu, M. (1996) Post-transcriptional reporter transgene silencing in transgenic tobacco. In *Mechanisms and Applications of Gene Silencing* (Grierson, D., Lycett, G.W. and Tucker, G.A. eds). Nottingham: Nottingham University Press, pp. 71–84.

Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A. and Sedat, J.W. (1996) Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell*, **85**, 745–759.

Dorer, D.R. and Henikoff, S. (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell*, **77**, 993–1002.

Dorlhac de Borne, F.D., Vincentz, M., Chupeau, Y. and Vaucheret, H. (1994) Co-suppression of nitrate reductase host genes and transgenes in transgenic tobacco plants. *Mol. Gen. Genet.* **243**, 613–621.

Dougherty, W.G., Lindbo, J.A., Smith, H.A., Parks, T.D., Swaney, S. and Proebsting, W.M. (1994) RNA-mediated virus resistance in transgenic plants: exploitation of a cellular pathway possibly involved in RNA degradation. *Mol. Plant–Microbe Interact.* **7**, 544–552.

Dougherty, W.G. and Parks, T.D. (1995) Transgenes and gene suppression: telling us something new? *Curr. Opin. Cell Biol.* **7**, 399–405.

Elmayan, T. and Vaucheret, H. (1996) A strongly-expressed 35S-driven transgene undergoes post-transcriptional silencing in all tobacco transformants irrespective of the copy number. *Plant J.* **9**, 787–797.

English, J.J., Mueller, E. and Baulcombe, D.C. (1996) Suppression of virus accumulation in transgenic plants exhibiting silencing of nuclear genes. *Plant Cell*, **8**, 179–188.

Fransz, P.F., Stam, M., Montijn, B., Ten Hoopen, R., Wiegant, J., Kooter, J.M., Oud, O. and Nanninga, N. (1996) Detection of single-copy genes and chromosome rearrangements in *Petunia hybrida* by fluorescence *in situ* hybridization. *Plant J.* **9**, 767–774.

Fray, R.G., Wallace, A.D. and Grierson, D. (1994) Identification of unexplained DNA fragments within the T-DNA borders of the Bin19 plant transformation vector. *Plant Mol. Biol.* **25**, 339–342.

Frisch, D.A., Harrishaller, L.W., Yokubaitis, N.T., Thomas, T.L., Hardin, S.H. and Hall, T.C. (1995) Complete sequence of the binary vector Bin19. *Plant Mol. Biol.* **27**, 405–409.

Goodwin, J., Chapman, K., Swaney, S., Parks, T.D., Wernsman, E.A. and Dougherty, W.G. (1996) Genetic and biochemical dissection of transgenic RNA-mediated virus resistance. *Plant Cell*, **8**, 95–105.

Gordenin, D.A., Lobachev, K.S., Degtyareva, N.P., Malkova, A.L., Perkins, E. and Resnick, M.A. (1993) Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol. Cell. Biol.* **13**, 5315–5322.

Hart, C.M., Fischer, B., Neuhaus, J.M. and Meins, F. (1992) Regulated inactivation of homologous gene expression in transgenic *Nicotiana sylvestris* plants containing a defense-related tobacco chitinase gene. *Mol. Gen. Genet.* **235**, 179–188.

Henderson, S.T. and Petes, T.D. (1993) Instability of a plasmid-borne inverted repeat in *Saccharomyces cerevisiae*. *Genetics*, **134**, 57–62.

Hobbs, S.L.A., Warkentin, T.D. and Delong, C.M.O. (1993) Transgene copy number can be positively or negatively associated with transgene expression. *Plant Mol. Biol.* **21**, 17–26.

Huala, E. and Sussex, I.M. (1993) Determination and cell interactions in reproductive meristems. *Plant Cell*, **5**, 1157–1165.

Ingelbrecht, I., Van Houdt, H., Van Montagu, M. and Depicker, A. (1994) Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl Acad. Sci. USA*, **91**, 10502–10506.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **6**, 3901–3907.

Jorgensen, R. (1992) Silencing of plant genes by homologous transgenes. *Agbiotech. News Info.* **4**, 265–273.

Jorgensen, R.A. (1995) Cosuppression, flower color patterns, and metastable gene expression states. *Science*, **268**, 686–691.

Jorgensen, R., Cluster, P., Que, Q., English, J. and Napoli, C. (1996) Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs. antisense constructs and single copy vs. complex T-DNA sequences. *Plant Mol. Biol.* **31**, 957–973.

Klimyuk, V.I., Carroll, B.J., Thomas, C.M. and Jones, J.D.G. (1993) Alkali treatment for rapid preparation of plant material for reliable PCR analysis. *Plant J.* **3**, 493–494.

Kunz, C., Schöb, H., Stam, M., Kooter, J.M. and Meins, F. (1996) Developmentally regulated silencing and reactivation of tobacco chitinase genes. *Plant J.* **10**, 437–450.

Leach, D.R.F. (1994) Long palindromes, cruciform structures, genetic instability and secondary structure repair. *BioEssays*, **16**, 893–900.

Lindbo, J.A., Silva-Rosales, L., Proebsting, W.M. and Dougherty, W.G. (1993) Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. *Plant Cell*, **5**, 1749–1759.

Martin, C. and Gerats, T. (1993) Control of pigment biosynthesis genes during petal development. *Plant Cell*, **5**, 1253–1264.

Martineau, B., Voelker, T.A. and Sanders, R.A. (1994) On defining T-DNA. *Plant Cell*, **11**, 1032–1033.

Matzke, A.J.M., Neuhuber, F., Park, Y.D., Ambros, P.F. and Matzke, M.A. (1994) Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes. *Mol. Gen. Genet.* **244**, 219–229.

Matzke, M.A. and Matzke, A.J.M. (1995) How and why do plants inactivate homologous (trans)genes? *Plant Physiol.* **107**, 679–685.

Matzke, M.A., Primig, M., Trnovsky, J. and Matzke, A.J.M. (1989)

Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* **8**, 643–649.

Metzlaff, M., O'Dell, M. and Flavell, R.B. (1996) Suppression of chalcone synthase A activities in *Petunia* by the addition of a transgene encoding chalcone synthase A. In *Mechanisms and Applications of Gene Silencing* (Grierson, D., Lycett, G.W. and Tucker, G.A., eds). Nottingham: Nottingham University Press, pp. 21–32.

Meyer, P. (1995) Understanding and controlling transgene expression. *Trends Biotechnol.* **13**, 332–337.

Meyer, P. (1996) Repeat-induced gene silencing: common mechanisms in plants and fungi. *Biol. Chem. H-S* **377**, 87–95.

Meyer, P., Heidmann, I. and Niedenhof, I. (1993) Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. *Plant J.* **4**, 89–100.

Mueller, E., Gilbert, J., Davenport, G., Brigneti, G. and Baulcombe, D.C. (1995) Homology-dependent resistance: transgenic virus resistance in plants related to homology-dependent gene silencing. *Plant J.* **7**, 1001–1013.

Napoli, C., Lemieux, C. and Jorgensen, R. (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*. *Plant Cell*, **2**, 279–289.

Neuhuber, F., Park, Y.D., Matzke, A.J.M. and Matzke, M.A. (1994) Susceptibility of transgene loci to homology-dependent gene silencing. *Mol. Gen. Genet.* **244**, 230–241.

Palauqui, J.-P. and Vaucheret, H. (1995) Field trial analysis of nitrate reductase co-suppression: a comparative study of 38 combinations of transgene loci. *Plant Mol. Biol.* **29**, 149–159.

Park, Y.-D., Papp, I., Moscone, E.A., Iglesias, V.A., Vaucheret, H., Matzke, A.J.M. and Matzke, M.A. (1996) Gene silencing mediated by promoter homology occurs at the level of transcription and results in meiotically heritable alterations in methylation and gene activity. *Plant J.* **9**, 183–194.

Ruskin, B. and Fink, G.R. (1993) Mutations in *POL1* increase the mitotic instability of tandem inverted repeats in *Saccharomyces cerevisiae*. *Genetics*, **134**, 43–56.

Smith, H.A., Swaney, S.L., Parks, T.D., Wernsman, E.A. and Dougherty, W.G. (1994) Transgenic plant virus resistance mediated by untranslatable sense RNAs: expression, regulation, and fate of nonessential RNAs. *Plant Cell*, **6**, 1441–1453.

Stam, M., Mol, J.N.M. and Kooter, J.M. (1997) The silence of genes in transgenic plants. *Ann. Bot.* (in press).

Van Blokland, R. (1994) Trans-inactivation of flower pigmentation genes in *Petunia hybrida*. PhD. Thesis. Amsterdam, The Netherlands: Vrije Universiteit

Van Blokland, R., Van der Geest, N., Mol, J.N.M. and Kooter, J.M. (1994) Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. *Plant J.* **6**, 861–877.

Van Blokland, R., Van der Geest, N., De Lange, P., Stam, M., Mol, J.N.M. and Kooter, J.M. (1996) Post-transcriptional suppression of chalcone synthase genes in *Petunia hybrida* and the accumulation of unspliced pre-mRNAs. In *Mechanisms and Applications of Gene Silencing* (Grierson, D., Lycett, G.W. and Tucker, G.A., eds). Nottingham: Nottingham University Press, pp. 57–69.

Van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N.M. and Stuitje, A.R. (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell*, **2**, 291–299.

Van Tunen, A.J., Mur, L., Brouns, G.S., Rienstra, J.-D., Koes, R.E. and Mol, J.N.M. (1990) Pollen- and anther-specific chi promoters from petunia: tandem promoter regulation of the *Chi* A gene. *Plant Cell*, **2**, 393–401.

Vaucheret, H. (1993) Identification of a general silencer for 19S and 35S promoters in a transgenic tobacco plant: 90 bp of homology in the promoter sequence are sufficient for trans-inactivation. *C. R. Acad. Sci. [III]*, **316**, 1471–1483.

Vaucheret, H., Palauqui, J.-C., Elmayan, T. and Moffatt, B. (1995) Molecular and genetic analysis of nitrite reductase co-suppression in transgenic tobacco plants. *Mol. Gen. Genet.* **248**, 311–317.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/11, 15/63, 15/82 | A1

(11) International Publication Number: **WO 99/32619**

(43) International Publication Date: 1 July 1999 (01.07.99)

(21) International Application Number: PCT/US98/27233

(22) International Filing Date: 21 December 1998 (21.12.98)

(30) Priority Data:

60/068,562	23 December 1997 (23.12.97)	US
09/215,257	18 December 1998 (18.12.98)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	60/068,562 (CON)
Filed on	23 December 1997 (23.12.97)
US	09/215,257 (CON)
Filed on	18 December 1998 (18.12.98)

(71) Applicants *(for all designated States except US)*: THE CARNEGIE INSTITUTE OF WASHINGTON [US/US]; 1530 P Street, N.W., Washington, DC 20005 (US). THE UNIVERSITY OF MASSACHUSETTS [US/US]; One Beacon Street, Boston, MA 02108 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: FIRE, Andrew [US/US]; 2320 Bright Leaf Way, Baltimore, MD 21210 (US). XU, Siqun [CN/US]; 1755 Warminton Court, Ballwin, MO 63021 (US). MONTGOMERY, Mary, K. [US/US]; 233 Macalester Street, St. Paul, MN 55105 (US). KOSTAS, Stephen, A. [US/US]; 126 East Melrose Avenue, Baltimore, MD 21212 (US). TIMMONS, Lisa [US/US]; 2408 Brambleton Road, Baltimore, MD 21209 (US). TABARA, Hiroaki [JP/US]; Apartment #1, 145 Orient Street, Worcester, MA 01604 (US). DRIVER, Samuel, E. [US/US]; Apartment #4, 1714 Commonwealth Avenue, Brighton, MA 02135 (US). MELLO, Craig, C. [US/US]; 19 Ryan Road, Shrewsbury, MA 01545 (US).

(74) Agents: KOKULIS, Paul, N. et al.; Pillsbury Madison & Sutro LLP, 1100 New York Avenue, N.W., Washington, DC 20005 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GENETIC INHIBITION BY DOUBLE-STRANDED RNA

(57) Abstract

A process is provided of introducing an RNA into a living cell to inhibit gene expression of a target gene in that cell. The process may be practiced *ex vivo* or *in vivo*. The RNA has a region with double-stranded structure. Inhibition is sequence-specific in that the nucleotide sequences of the duplex region of the RNA and of a portion of the target gene are identical. The present invention is distinguished from prior art interference in gene expression by antisense or triple-strand methods.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco.	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GENETIC INHIBITION BY DOUBLE-STRANDED RNA

GOVERNMENT RIGHTS

This invention was made with U.S. government support under grant numbers GM-37706, GM-17164, HD-33769 and GM-07231 awarded by the National Institutes of Health. The U.S. government has certain rights in the invention.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to gene-specific inhibition of gene expression by double-stranded ribonucleic acid (dsRNA).

2. Description of the Related Art

Targeted inhibition of gene expression has been a long-felt need in biotechnology and genetic engineering. Although a major investment of effort has been made to achieve this goal, a more comprehensive solution to this problem was still needed.

Classical genetic techniques have been used to isolate mutant organisms with reduced expression of selected genes. Although valuable, such techniques require laborious mutagenesis and screening programs, are limited to organisms in which genetic manipulation is well established (e.g., the existence of selectable markers, the ability to control genetic segregation and sexual reproduction), and are limited to applications in which a large number of cells or organisms can be sacrificed to isolate the desired mutation. Even under these circumstances, classical genetic techniques can fail to produce mutations in specific target genes of interest, particularly when complex genetic pathways are involved. Many applications of molecular genetics require the ability to go beyond classical genetic screening techniques and efficiently produce a *directed* change in gene expression in a specified group of cells or organisms. Some such applications are knowledge-based projects in which it is of importance to understand what effects the loss of a specific gene product (or products) will have on the behavior of the cell or organism. Other applications are engineering based, for example: cases in which is important to

produce a population of cells or organisms in which a specific gene product (or products) has been reduced or removed. A further class of applications is therapeutically based in which it would be valuable for a functioning organism (e.g., a human) to reduce or remove the amount of a specified gene product (or products). Another class of applications provides a disease model in which a physiological function in a living organism is genetically manipulated to reduce or remove a specific gene product (or products) without making a permanent change in the organism's genome.

In the last few years, advances in nucleic acid chemistry and gene transfer have inspired new approaches to engineer specific interference with gene expression. These approaches are described below.

Use of Antisense Nucleic Acids to Engineer Interference

Antisense technology has been the most commonly described approach in protocols to achieve gene-specific interference. For antisense strategies, stochiometric amounts of single-stranded nucleic acid complementary to the messenger RNA for the gene of interest are introduced into the cell. Some difficulties with antisense-based approaches relate to delivery, stability, and dose requirements. In general, cells do not have an uptake mechanism for single-stranded nucleic acids, hence uptake of unmodified single-stranded material is extremely inefficient. While waiting for uptake into cells, the single-stranded material is subject to degradation. Because antisense interference requires that the interfering material accumulate at a relatively high concentration (at or above the concentration of endogenous mRNA), the amount required to be delivered is a major constraint on efficacy. As a consequence, much of the effort in developing antisense technology has been focused on the production of modified nucleic acids that are both stable to nuclease digestion and able to diffuse readily into cells. The use of antisense interference for gene therapy or other whole-organism applications has been limited by the large amounts of oligonucleotide that need to be synthesized from non-natural analogs, the cost of such synthesis, and the difficulty even with high doses of maintaining a sufficiently concentrated and uniform pool of interfering material in each cell.

Triple-Helix Approaches to Engineer Interference

A second, proposed method for engineered interference is based on a triple helical nucleic acid structure. This approach relies on the rare ability of certain nucleic acid populations to adopt a triple-stranded structure. Under physiological conditions, nucleic acids are virtually all single- or double-stranded, and rarely if ever form triple-stranded structures. It has been known for some time, however, that certain simple purine- or pyrimidine-rich sequences could form a triple-stranded molecule *in vitro* under extreme conditions of pH (i.e., in a test tube). Such structures are generally very transient under physiological conditions, so that simple delivery of unmodified nucleic acids designed to produce triple-strand structures does not yield interference. As with antisense, development of triple-strand technology for use *in vivo* has focused on the development of modified nucleic acids that would be more stable and more readily absorbed by cells *in vivo*. An additional goal in developing this technology has been to produce modified nucleic acids for which the formation of triple-stranded material proceeds effectively at physiological pH.

Co-Suppression Phenomena and Their Use in Genetic Engineering

A third approach to gene-specific interference is a set of operational procedures grouped under the name "co-suppression". This approach was first described in plants and refers to the ability of transgenes to cause silencing of an unlinked but homologous gene. More recently, phenomena similar to co-suppression have been reported in two animals: *C. elegans* and *Drosophila*. Co-suppression was first observed by accident, with reports coming from groups using transgenes in attempts to achieve over-expression of a potentially useful locus. In some cases the over-expression was successful while, in many others, the result was opposite from that expected. In those cases, the transgenic plants actually showed less expression of the endogenous gene. Several mechanisms have so far been proposed for transgene-mediated co-suppression in plants; all of these mechanistic proposals remain hypothetical, and no definitive mechanistic description of the process has been presented. The models that have been proposed to explain co-suppression can be placed in two different categories. In one set of proposals, a direct physical interaction at the DNA- or chromatin-level between two different chromosomal sites has been

hypothesized to occur; an as-yet-unidentified mechanism would then lead to *de novo* methylation and subsequent suppression of gene expression. Alternatively, some have postulated an RNA intermediate, synthesized at the transgene locus, which might then act to produce interference with the endogenous gene. The characteristics of the interfering RNA, as well as the nature of the interference process, have not been determined. Recently, a set of experiments with RNA viruses have provided some support for the possibility of RNA intermediates in the interference process. In these experiments, a replicating RNA virus is modified to include a segment from a gene of interest. This modified virus is then tested for its ability to interfere with expression of the endogenous gene. Initial results with this technique have been encouraging, however, the properties of the viral RNA that are responsible for interference effects have not been determined and, in any case, would be limited to plants which are hosts of the plant virus.

Distinction between the Present Invention and Antisense Approaches

The present invention differs from antisense-mediated interference in both approach and effectiveness. Antisense-mediated genetic interference methods have a major challenge: delivery to the cell interior of specific single-stranded nucleic acid molecules at a concentration that is equal to or greater than the concentration of endogenous mRNA. Double-stranded RNA-mediated inhibition has advantages both in the stability of the material to be delivered and the concentration required for effective inhibition. Below, we disclose that in the model organism *C. elegans*, the present invention is at least 100-fold more effective than an equivalent antisense approach (i.e., dsRNA is at least 100-fold more effective than the injection of purified antisense RNA in reducing gene expression). These comparisons also demonstrate that inhibition by double-stranded RNA must occur by a mechanism distinct from antisense interference.

Distinction between the Present Invention and Triple-Helix Approaches

The limited data on triple strand formation argues against the involvement of a stable triple-strand intermediate in the present invention. Triple-strand structures occur rarely, if at all, under physiological conditions and are limited to very unusual base sequence with long runs of purines and pyrimidines. By contrast, dsRNA-mediated

inhibition occurs efficiently under physiological conditions, and occurs with a wide variety of inhibitory and target nucleotide sequences. The present invention has been used to inhibit expression of 18 different genes, providing phenocopies of null mutations in these genes of known function. The extreme environmental and sequence constraints on triple-helix formation make it unlikely that dsRNA-mediated inhibition *in C. elegans* is mediated by a triple-strand structure.

Distinction between Present Invention and Co-Suppression Approaches

The transgene-mediated genetic interference phenomenon called co-suppression may include a wide variety of different processes. From the viewpoint of application to other types of organisms, the co-suppression phenomenon in plants is difficult to extend. A confounding aspect in creating a general technique based on co-suppression is that some transgenes in plants lead to suppression of the endogenous locus and some do not. Results in *C. elegans* and *Drosophila* indicate that certain transgenes can cause interference (i.e., a quantitative decrease in the activity of the corresponding endogenous locus) but that most transgenes do not produce such an effect. The lack of a predictable effect in plants, nematodes, and insects greatly limits the usefulness of simply adding transgenes to the genome to interfere with gene expression. Viral-mediated co-suppression in plants appears to be quite effective, but has a number of drawbacks. First, it is not clear what aspects of the viral structure are critical for the observed interference. Extension to another system would require discovery of a virus in that system which would have these properties, and such a library of useful viral agents are not available for many organisms. Second, the use of a replicating virus within an organism to effect genetic changes (e.g., long- or short-term gene therapy) requires considerably more monitoring and oversight for deleterious effects than the use of a defined nucleic acid as in the present invention.

The present invention avoids the disadvantages of the previously-described methods for genetic interference. Several advantages of the present invention are discussed below, but numerous others will be apparent to one of ordinary skill in the biotechnology and genetic engineering arts.

SUMMARY OF THE INVENTION

A process is provided for inhibiting expression of a target gene in a cell. The process comprises introduction of RNA with partial or fully double-stranded character into the cell or into the extracellular environment. Inhibition is specific in that a nucleotide sequence from a portion of the target gene is chosen to produce inhibitory RNA. We disclose that this process is (1) effective in producing inhibition of gene expression, (2) specific to the targeted gene, and (3) general in allowing inhibition of many different types of target gene.

The target gene may be a gene derived from the cell, an endogenous gene, a transgene, or a gene of a pathogen which is present in the cell after infection thereof. Depending on the particular target gene and the dose of double stranded RNA material delivered, the procedure may provide partial or complete loss of function for the target gene. A reduction or loss of gene expression in at least 99% of targeted cells has been shown. Lower doses of injected material and longer times after administration of dsRNA may result in inhibition in a smaller fraction of cells. Quantitation of gene expression in a cell may show similar amounts of inhibition at the level of accumulation of target mRNA or translation of target protein.

The RNA may comprise one or more strands of polymerized ribonucleotide; it may include modifications to either the phosphate-sugar backbone or the nucleoside. The double-stranded structure may be formed by a single self-complementary RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The RNA may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses of double-stranded material may yield more effective inhibition. Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition. RNA containing a nucleotide sequences identical to a portion of the target gene is preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may optimized by alignment algorithms known in the art and calculating the percent difference between the nucleotide sequences. Alternatively, the duplex region of the RNA may be

defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript.

The cell with the target gene may be derived from or contained in any organism (e.g., plant, animal, protozoan, virus, bacterium, or fungus). RNA may be synthesized either *in vivo* or *in vitro*. Endogenous RNA polymerase of the cell may mediate transcription *in vivo*, or cloned RNA polymerase can be used for transcription *in vivo* or in *vitro*. For transcription from a transgene *in vivo* or an expression construct, a regulatory region may be used to transcribe the RNA strand (or strands).

The RNA may be directly introduced into the cell (i.e., intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism, introduced orally, or may be introduced by bathing an organism in a solution containing RNA. Methods for oral introduction include direct mixing of RNA with food of the organism, as well as engineered approaches in which a species that is used as food is engineered to express an RNA, then fed to the organism to be affected. Physical methods of introducing nucleic acids include injection directly into the cell or extracellular injection into the organism of an RNA solution.

The advantages of the present invention include: the ease of introducing double-stranded RNA into cells, the low concentration of RNA which can be used, the stability of double-stranded RNA, and the effectiveness of the inhibition. The ability to use a low concentration of a naturally-occurring nucleic acid avoids several disadvantages of antisense interference. This invention is not limited to *in vitro* use or to specific sequence compositions, as are techniques based on triple-strand formation. And unlike antisense interference, triple-strand interference, and co-suppression, this invention does not suffer from being limited to a particular set of target genes, a particular portion of the target gene's nucleotide sequence, or a particular transgene or viral delivery method. These concerns have been a serious obstacle to designing general strategies according to the prior art for inhibiting gene expression of a target gene of interest.

Furthermore, genetic manipulation becomes possible in organisms that are not classical genetic models. Breeding and screening programs may be accelerated by the ability to rapidly assay the consequences of a specific, targeted gene disruption. Gene disruptions may be used to discover the function of the target gene, to produce disease

models in which the target gene are involved in causing or preventing a pathological condition, and to produce organisms with improved economic properties.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the genes used to study RNA-mediated genetic inhibition in *C. elegans*. Intron-exon structure for genes used to test RNA-mediated inhibition are shown (exons: filled boxes; introns: open boxes; 5' and 3' untranslated regions: shaded; *unc-22* [9], *unc-54* [12], *fem-1* [14], and *hlh-1* [15]).

Figures 2 A-I show analysis of inhibitory RNA effects in individual cells. These experiments were carried out in a reporter strain (called PD4251) expressing two different reporter proteins, nuclear GFP-LacZ and mitochondrial GFP. The micrographs show progeny of injected animals visualized by a fluorescence microscope. Panels A (young larva), B (adult), and C (adult body wall; high magnification) result from injection of a control RNA (ds-*unc22A*). Panels D-F show progeny of animals injected with ds-*gfpG*. Panels G-I demonstrate specificity. Animals are injected with ds-*lacZL* RNA, which should affect the nuclear but not the mitochondrial reporter construct. Panel H shows a typical adult, with nuclear GFP-LacZ lacking in almost all body-wall muscles but retained in vulval muscles. Scale bars are 20 μm.

Figures 3 A-D show effects of double-stranded RNA corresponding to *mex-3* on levels of the endogenous mRNA. Micrographs show *in situ* hybridization to embryos (dark stain). Panel A: Negative control showing lack of staining in the absence of hybridization probe. Panel B: Embryo from uninjected parent (normal pattern of endogenous *mex-3* RNA[20]). Panel C: Embryo from a parent injected with purified *mex-3*B antisense RNA. These embryos and the parent animals retain the *mex-3* mRNA, although levels may have been somewhat less than wild type. Panel D: Embryo from a parent injected with dsRNA corresponding to *mex-3*B; no *mex-3* RNA was detected. Scale: each embryo is approximately 50 μm in length.

Figure 4 shows inhibitory activity of *unc-22A* as a function of structure and concentration. The main graph indicates fractions in each behavioral class. Embryos in the uterus and already covered with an eggshell at the time of injection were not affected and, thus, are not included. Progeny cohort groups are labeled 1 for 0-6 hours, 2 for 6-15

hours, 3 for 15-27 hours, 4 for 27-41 hours, and 5 for 41-56 hours The bottom-left diagram shows genetically derived relationship between *unc-22* gene dosage and behavior based on analyses of *unc-22* heterozygotes and polyploids[8,3].

Figures 5 A-C show examples of genetic inhibition following ingestion by *C. elegans* of dsRNAs from expressing bacteria. Panel A: General strategy for production of dsRNA by cloning a segment of interest between flanking copies of the bacteriophage T7 promoter and transcribing both strands of the segment by transfecting a bacterial strain (BL21/DE3)[28] expressing the T7 polymerase gene from an inducible (Lac) promoter. Panel B: A GFP-expressing *C. elegans* strain, PD4251 (see Figure 2), fed on a native bacterial host. Panel C: PD4251 animals reared on a diet of bacteria expressing dsRNA corresponding to the coding region for *gfp*.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method of producing sequence-specific inhibition of gene expression by introducing double-stranded RNA (dsRNA). A process is provided for inhibiting expression of a target gene in a cell. The process comprises introduction of RNA with partial or fully double-stranded character into the cell. Inhibition is sequence-specific in that a nucleotide sequence from a portion of the target gene is chosen to produce inhibitory RNA. We disclose that this process is (1) effective in producing inhibition of gene expression, (2) specific to the targeted gene, and (3) general in allowing inhibition of many different types of target gene.

The target gene may be a gene derived from the cell (i.e., a cellular gene), an endogenous gene (i.e., a cellular gene present in the genome), a transgene (i.e., a gene construct inserted at an ectopic site in the genome of the cell), or a gene from a pathogen which is capable of infecting an organism from which the cell is derived. Depending on the particular target gene and the dose of double stranded RNA material delivered, this process may provide partial or complete loss of function for the target gene. A reduction or loss of gene expression in at least 99% of targeted cells has been shown.

Inhibition of gene expression refers to the absence (or observable decrease) in the level of protein and/or mRNA product from a target gene. Specificity refers to the ability to inhibit the target gene without manifest effects on other genes of the cell. The

consequences of inhibition can be confirmed by examination of the outward properties of the cell or organism (as presented below in the examples) or by biochemical techniques such as RNA solution hybridization, nuclease protection, Northern hybridization, reverse transcription, gene expression monitoring with a microarray, antibody binding, enzyme linked immunosorbent assay (ELISA), Western blotting, radioimmunoassay (RIA), other immunoassays, and fluorescence activated cell analysis (FACS). For RNA-mediated inhibition in a cell line or whole organism, gene expression is conveniently assayed by use of a reporter or drug resistance gene whose protein product is easily assayed. Such reporter genes include acetohydroxyacid synthase (AHAS), alkaline phosphatase (AP), beta galactosidase (LacZ), beta glucoronidase (GUS), chloramphenicol acetyltransferase (CAT), green fluorescent protein (GFP), horseradish peroxidase (HRP), luciferase (Luc), nopaline synthase (NOS), octopine synthase (OCS), and derivatives thereof. Multiple selectable markers are available that confer resistance to ampicillin, bleomycin, chloramphenicol, gentamycin, hygromycin, kanamycin, lincomycin, methotrexate, phosphinothricin, puromycin, and tetracyclin.

Depending on the assay, quantitation of the amount of gene expression allows one to determine a degree of inhibition which is greater than 10%, 33%, 50%, 90%, 95% or 99% as compared to a cell not treated according to the present invention. Lower doses of injected material and longer times after administration of dsRNA may result in inhibition in a smaller fraction of cells (e.g., at least 10%, 20%, 50%, 75%, 90%, or 95% of targeted cells). Quantitation of gene expression in a cell may show similar amounts of inhibition at the level of accumulation of target mRNA or translation of target protein. As an example, the efficiency of inhibition may be determined by assessing the amount of gene product in the cell: mRNA may be detected with a hybridization probe having a nucleotide sequence outside the region used for the inhibitory double-stranded RNA, or translated polypeptide may be detected with an antibody raised against the polypeptide sequence of that region.

The RNA may comprise one or more strands of polymerized ribonucleotide. It may include modifications to either the phosphate-sugar backbone or the nucleoside. For example, the phosphodiester linkages of natural RNA may be modified to include at least one of a nitrogen or sulfur heteroatom. Modifications in RNA structure may be tailored

to allow specific genetic inhibition while avoiding a general panic response in some organisms which is generated by dsRNA. Likewise, bases may be modified to block the activity of adenosine deaminase. RNA may be produced enzymatically or by partial/total organic synthesis, any modified ribonucleotide can be introduced by *in vitro* enzymatic or organic synthesis.

The double-stranded structure may be formed by a single self-complementary RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The RNA may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses (e.g., at least 5, 10, 100, 500 or 1000 copies per cell) of double-stranded material may yield more effective inhibition; lower doses may also be useful for specific applications. Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition.

RNA containing a nucleotide sequences identical to a portion of the target gene are preferred for inhibition. RNA sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, *Sequence Analysis Primer*, Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory RNA and the portion of the target gene is preferred. Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript (e.g., 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 50°C or 70°C hybridization for 12-16 hours; followed by washing). The length of the identical nucleotide sequences may be at least 25. 50, 100, 200, 300 or 400 bases.

As disclosed herein, 100% sequence identity between the RNA and the target gene is not required to practice the present invention. Thus the invention has the advantage of

being able to tolerate sequence variations that might be expected due to genetic mutation, strain polymorphism, or evolutionary divergence.

The cell with the target gene may be derived from or contained in any organism. The organism may a plant, animal, protozoan, bacterium, virus, or fungus. The plant may be a monocot, dicot or gymnosperm; the animal may be a vertebrate or invertebrate. Preferred microbes are those used in agriculture or by industry, and those that are pathogenic for plants or animals. Fungi include organisms in both the mold and yeast morphologies.

Plants include arabidopsis; field crops (e.g., alfalfa, barley, bean, corn, cotton, flax, pea, rape, rice, rye, safflower, sorghum, soybean, sunflower, tobacco, and wheat); vegetable crops (e.g., asparagus, beet, broccoli, cabbage, carrot, cauliflower, celery, cucumber, eggplant, lettuce, onion, pepper, potato, pumpkin, radish, spinach, squash, taro, tomato, and zucchini); fruit and nut crops (e.g., almond, apple, apricot, banana, blackberry, blueberry, cacao, cherry, coconut, cranberry, date, fajoa, filbert, grape, grapefruit, guava, kiwi, lemon, lime, mango, melon, nectarine, orange, papaya, passion fruit, peach, peanut, pear, pineapple, pistachio, plum, raspberry, strawberry, tangerine, walnut, and watermelon); and ornamentals (e.g., alder, ash, aspen, azalea, birch, boxwood, camellia, carnation, chrysanthemum, elm, fir, ivy, jasmine, juniper, oak, palm, poplar, pine, redwood, rhododendron, rose, and rubber).

Examples of vertebrate animals include fish, mammal, cattle, goat, pig, sheep, rodent, hamster, mouse, rat, primate, and human; invertebrate animals include nematodes, other worms, drosophila, and other insects. Representative generae of nematodes include those that infect animals (e.g., Ancylostoma, Ascaridia, Ascaris, Bunostomum, Caenorhabditis, Capillaria, Chabertia, Cooperia, Dictyocaulus, Haemonchus, Heterakis, Nematodirus, Oesophagostomum, Ostertagia, Oxyuris, Parascaris, Strongylus, Toxascaris, Trichuris, Trichostrongylus, Tfhchonema, Toxocara, Uncinaria) and those that infect plants (e.g., Bursaphalenchus, Criconemella, Diiylenchus, Ditylenchus, Globodera, Helicotylenchus, Heterodera, Longidorus, Melodoigyne, Nacobbus, Paratylenchus, Pratylenchus, Radopholus, Rotelynchus, Tylenchus, and Xiphinema). Representative orders of insects include Coleoptera, Diptera, Lepidoptera, and Homoptera.

The cell having the target gene may be from the germ line or somatic, totipotent or pluripotent, dividing or non-dividing, parenchyma or epithelium, immortalized or transformed, or the like. The cell may be a stem cell or a differentiated cell. Cell types that are differentiated include adipocytes, fibroblasts, myocytes, cardiomyocytes, endothelium, neurons, glia, blood cells, megakaryocytes, lymphocytes, macrophages, neutrophils, eosinophils, basophils, mast cells, leukocytes, granulocytes, keratinocytes, chondrocytes, osteoblasts, osteoclasts, hepatocytes, and cells of the endocrine or exocrine glands.

RNA may be synthesized either *in vivo* or *in vitro*. Endogenous RNA polymerase of the cell may mediate transcription *in vivo*, or cloned RNA polymerase can be used for transcription *in vivo* or in *vitro*. For transcription from a transgene *in vivo* or an expression construct, a regulatory region (e.g., promoter, enhancer, silencer, splice donor and acceptor, polyadenylation) may be used to transcribe the RNA strand (or strands). Inhibition may be targeted by specific transcription in an organ, tissue, or cell type; stimulation of an environmental condition (e.g., infection, stress, temperature, chemical inducers); and/or engineering transcription at a developmental stage or age. The RNA strands may or may not be polyadenylated; the RNA strands may or may not be capable of being translated into a polypeptide by a cell's translational apparatus. RNA may be chemically or enzymatically synthesized by manual or automated reactions. The RNA may be synthesized by a cellular RNA polymerase or a bacteriophage RNA polymerase (e.g., T3, T7, SP6). The use and production of an expression construct are known in the art[32, 33, 34] (see also WO 97/32016; U.S. Pat. Nos. 5,593,874, 5,698,425, 5,712,135, 5,789,214, and 5,804,693; and the references cited therein). If synthesized chemically or by *in vitro* enzymatic synthesis, the RNA may be purified prior to introduction into the cell. For example, RNA can be purified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography, or a combination thereof. Alternatively, the RNA may be used with no or a minimum of purification to avoid losses due to sample processing. The RNA may be dried for storage or dissolved in an aqueous solution. The solution may contain buffers or salts to promote annealing, and/or stabilization of the duplex strands.

RNA may be directly introduced into the cell (i.e., intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism, intro-

duced orally, or may be introduced by bathing an organism in a solution containing the RNA. Methods for oral introduction include direct mixing of the RNA with food of the organism, as well as engineered approaches in which a species that is used as food is engineered to express the RNA, then fed to the organism to be affected. For example, the RNA may be sprayed onto a plant or a plant may be genetically engineered to express the RNA in an amount sufficient to kill some or all of a pathogen known to infect the plant. Physical methods of introducing nucleic acids, for example, injection directly into the cell or extracellular injection into the organism, may also be used. We disclose herein that in *C. elegans*, double-stranded RNA introduced outside the cell inhibits gene expression. Vascular or extravascular circulation, the blood or lymph system, the phloem, the roots, and the cerebrospinal fluid are sites where the RNA may be introduced. A transgenic organism that expresses RNA from a recombinant construct may be produced by introducing the construct into a zygote, an embryonic stem cell, or another multipotent cell derived from the appropriate organism.

Physical methods of introducing nucleic acids include injection of a solution containing the RNA, bombardment by particles covered by the RNA, soaking the cell or organism in a solution of the RNA, or electroporation of cell membranes in the presence of the RNA. A viral construct packaged into a viral particle would accomplish both efficient introduction of an expression construct into the cell and transcription of RNA encoded by the expression construct. Other methods known in the art for introducing nucleic acids to cells may be used, such as lipid-mediated carrier transport, chemical-mediated transport, such as calcium phosphate, and the like. Thus the RNA may be introduced along with components that perform one or more of the following activities: enhance RNA uptake by the cell, promote annealing of the duplex strands, stabilize the annealed strands, or other-wise increase inhibition of the target gene.

The present invention may be used to introduce RNA into a cell for the treatment or prevention of disease. For example, dsRNA may be introduced into a cancerous cell or tumor and thereby inhibit gene expression of a gene required for maintenance of the carcinogenic/tumorigenic phenotype. To prevent a disease or other pathology, a target gene may be selected which is required for initiation or maintenance of the disease/pathology.

Treatment would include amelioration of any symptom associated with the disease or clinical indication associated with the pathology.

A gene derived from any pathogen may be targeted for inhibition. For example, the gene could cause immunosuppression of the host directly or be essential for replication of the pathogen, transmission of the pathogen, or maintenance of the infection. The inhibitory RNA could be introduced in cells *in vitro* or *ex vivo* and then subsequently placed into an animal to affect therapy, or directly treated by *in vivo* administration. A method of gene therapy can be envisioned. For example, cells at risk for infection by a pathogen or already infected cells, particularly human immunodeficiency virus (HIV) infections, may be targeted for treatment by introduction of RNA according to the invention. The target gene might be a pathogen or host gene responsible for entry of a pathogen into its host, drug metabolism by the pathogen or host, replication or integration of the pathogen's genome, establishment or spread of an infection in the host, or assembly of the next generation of pathogen. Methods of prophylaxis (i.e., prevention or decreased risk of infection), as well as reduction in the frequency or severity of symptoms associated with infection, can be envisioned.

The present invention could be used for treatment or development of treatments for cancers of any type, including solid tumors and leukemias, including: apudoma, choristoma, branchioma, malignant carcinoid syndrome, carcinoid heart disease, carcinoma (e.g., Walker, basal cell, basosquamous, Brown-Pearce, ductal, Ehrlich tumor, in situ, Krebs 2, Merkel cell, mucinous, non-small cell lung, oat cell, papillary, scirrhous, bronchiolar, bronchogenic, squamous cell, and transitional cell), histiocytic disorders, leukemia (e.g., B cell, mixed cell, null cell, T cell, T-cell chronic, HTLV-II-associated, lymphocytic acute, lymphocytic chronic, mast cell, and myeloid), histiocytosis malignant, Hodgkin disease, immunoproliferative small, non-Hodgkin lymphoma, plasmacytoma, reticuloendotheliosis, melanoma, chondroblastoma, chondroma, chondrosarcoma, fibroma, fibrosarcoma, giant cell tumors, histiocytoma, lipoma, liposarcoma, mesothelioma, myxoma, myxosarcoma, osteoma, osteosarcoma, Ewing sarcoma, synovioma, adenofibroma, adenolymphoma, carcinosarcoma, chordoma, cranio-pharyngioma, dysgerminoma, hamartoma, mesenchymoma, mesonephroma, myosarcoma, ameloblastoma, cementoma, odontoma, teratoma, thymoma, trophoblastic tumor, adeno-

carcinoma, adenoma, cholangioma, cholesteatoma, cylindroma, cystadenocarcinoma, cystadenoma, granulosa cell tumor, gynandroblastoma, hepatoma, hidradenoma, islet cell tumor, Leydig cell tumor, papilloma, Sertoli cell tumor, theca cell tumor, leiomyoma, leiomyosarcoma, myoblastoma, myoma, myosarcoma, rhabdomyoma, rhabdomyosarcoma, ependymoma, ganglioneuroma, glioma, medulloblastoma, meningioma, neurilemmoma, neuroblastoma, neuroepithelioma, neurofibroma, neuroma, paraganglioma, paraganglioma nonchromaffin, angiokeratoma, angiolymphoid hyperplasia with eosinophilia, angioma sclerosing, angiomatosis, glomangioma, hemangioendothelioma, hemangioma, hemangiopericytoma, hemangiosarcoma, lymphangioma, lymphangiomyoma, lymphangiosarcoma, pinealoma, carcinosarcoma, chondrosarcoma, cystosarcoma phyllodes, fibrosarcoma, hemangiosarcoma, leiomyosarcoma, leukosarcoma, liposarcoma, lymphangiosarcoma, myosarcoma, myxosarcoma, ovarian carcinoma, rhabdomyosarcoma, sarcoma (e.g., Ewing, experimental, Kaposi, and mast cell), neoplasms (e.g., bone, breast, digestive system, colorectal, liver, pancreatic, pituitary, testicular, orbital, head and neck, central nervous system, acoustic, pelvic, respiratory tract, and urogenital), neurofibromatosis, and cervical dysplasia, and for treatment of other conditions in which cells have become immortalized or transformed. The invention could be used in combination with other treatment modalities, such as chemotherapy, cryotherapy, hyperthermia, radiation therapy, and the like.

As disclosed herein, the present invention may is not limited to any type of target gene or nucleotide sequence. But the following classes of possible target genes are listed for illustrative purposes: developmental genes (e.g., adhesion molecules, cyclin kinase inhibitors, Wnt family members, Pax family members, Winged helix family members, Hox family members, cytokines/lymphokines and their receptors, growth/differentiation factors and their receptors, neurotransmitters and their receptors); oncogenes (e.g., ABL1, BCL1, BCL2, BCL6, CBFA2, CBL, CSF1R, ERBA, ERBB, EBRB2, ETS1, ETS1, ETV6, FGR, FOS, FYN, HCR, HRAS, JUN, KRAS, LCK, LYN, MDM2, MLL, MYB, MYC, MYCL1, MYCN, NRAS, PIM1, PML, RET, SRC, TAL1, TCL3, and YES); tumor suppressor genes (e.g., APC, BRCA1, BRCA2, MADH4, MCC, NF1, NF2, RB1, TP53, and WT1); and enzymes (e.g., ACC synthases and oxidases, ACP desaturases and hydroxylases, ADP-glucose pyrophorylases, ATPases, alcohol dehydrogenases, amylases,

amyloglucosidases, catalases, cellulases, chalcone synthases, chitinases, cyclooxygenases, decarboxylases, dextrinases, DNA and RNA polymerases, galactosidases, glucanases, glucose oxidases, granule-bound starch synthases, GTPases, helicases, hemicellulases, integrases, inulinases, invertases, isomerases, kinases, lactases, lipases, lipoxygenases, lysozymes, nopaline synthases, octopine synthases, pectinesterases, peroxidases, phosphatases, phospholipases, phosphorylases, phytases, plant growth regulator synthases, polygalacturonases, proteinases and peptidases, pullanases, recombinases, reverse transcriptases, RUBISCOs, topoisomerases, and xylanases).

The present invention could comprise a method for producing plants with reduced susceptibility to climatic injury, susceptibility to insect damage, susceptibility to infection by a pathogen, or altered fruit ripening characteristics. The targeted gene may be an enzyme, a plant structural protein, a gene involved in pathogenesis, or an enzyme that is involved in the production of a non-proteinaceous part of the plant (i.e., a carbohydrate or lipid). If an expression construct is used to transcribe the RNA in a plant, transcription by a wound- or stress-inducible; tissue-specific (e.g., fruit, seed, anther, flower, leaf, root); or otherwise regulatable (e.g., infection, light, temperature, chemical) promoter may be used. By inhibiting enzymes at one or more points in a metabolic pathway or genes involved in pathogenesis, the effect may be enhanced: each activity will be affected and the effects may be magnified by targeting multiple different components. Metabolism may also be manipulated by inhibiting feedback control in the pathway or production of unwanted metabolic byproducts.

The present invention may be used to reduce crop destruction by other plant pathogens such as arachnids, insects, nematodes, protozoans, bacteria, or fungi. Some such plants and their pathogens are listed in *Index of Plant Diseases in the United States* (U.S. Dept. of Agriculture Handbook No. 165, 1960); *Distribution of Plant-Parasitic Nematode Species in North America* (Society of Nematologists, 1985); and *Fungi on Plants and Plant Products in the United States* (American Phytopathological Society, 1989). Insects with reduced ability to damage crops or improved ability to prevent other destructive insects from damaging crops may be produced. Furthermore, some nematodes are vectors of plant pathogens, and may be attacked by other beneficial nematodes which have no effect on plants. Inhibition of target gene activity could be used to delay or

prevent entry into a particular developmental step (e.g., metamorphosis), if plant disease was associated with a particular stage of the pathogen's life cycle. Interactions between pathogens may also be modified by the invention to limit crop damage. For example, the ability of beneficial nematodes to attack their harmful prey may be enhanced by inhibition of behavior-controlling nematode genes according to the invention.

Although pathogens cause disease, some of the microbes interact with their plant host in a beneficial manner. For example, some bacteria are involved in symbiotic relationships that fix nitrogen and some fungi produce phytohormones. Such beneficial interactions may be promoted by using the present invention to inhibit target gene activity in the plant and/or the microbe.

Another utility of the present invention could be a method of identifying gene function in an organism comprising the use of double-stranded RNA to inhibit the activity of a target gene of previously unknown function. Instead of the time consuming and laborious isolation of mutants by traditional genetic screening, functional genomics would envision determining the function of uncharacterized genes by employing the invention to reduce the amount and/or alter the timing of target gene activity. The invention could be used in determining potential targets for pharmaceutics, understanding normal and pathological events associated with development, determining signaling pathways responsible for postnatal development/aging, and the like. The increasing speed of acquiring nucleotide sequence information from genomic and expressed gene sources, including total sequences for the yeast, *D. melanogaster*, and *C. elegans* genomes, can be coupled with the invention to determine gene function in an organism (e.g., nematode). The preference of different organisms to use particular codons, searching sequence databases for related gene products, correlating the linkage map of genetic traits with the physical map from which the nucleotide sequences are derived, and artificial intelligence methods may be used to define putative open reading frames from the nucleotide sequences acquired in such sequencing projects.

A simple assay would be to inhibit gene expression according to the partial sequence available from an expressed sequence tag (EST). Functional alterations in growth, development, metabolism, disease resistance, or other biological processes would be indicative of the normal role of the EST's gene product.

The ease with which RNA can be introduced into an intact cell/organism containing the target gene allows the present invention to be used in high throughput screening (HTS). For example, duplex RNA can be produced by an amplification reaction using primers flanking the inserts of any gene library derived from the target cell/organism. Inserts may be derived from genomic DNA or mRNA (e.g., cDNA and cRNA). Individual clones from the library can be replicated and then isolated in separate reactions, but preferably the library is maintained in individual reaction vessels (e.g., a 96-well microtiter plate) to minimize the number of steps required to practice the invention and to allow automation of the process. Solutions containing duplex RNAs that are capable of inhibiting the different expressed genes can be placed into individual wells positioned on a microtiter plate as an ordered array, and intact cells/organisms in each well can be assayed for any changes or modifications in behavior or development due to inhibition of target gene activity. The amplified RNA can be fed directly to, injected into, the cell/organism containing the target gene. Alternatively, the duplex RNA can be produced by *in vivo* or *in vitro* transcription from an expression construct used to produce the library. The construct can be replicated as individual clones of the library and transcribed to produce the RNA; each clone can then be fed to, or injected into, the cell/organism containing the target gene. The function of the target gene can be assayed from the effects it has on the cell/organism when gene activity is inhibited. This screening could be amenable to small subjects that can be processed in large number, for example: arabidopsis, bacteria, drosophila, fungi, nematodes, viruses, zebrafish, and tissue culture cells derived from mammals.

A nematode or other organism that produces a colorimetric, fluorogenic, or luminescent signal in response to a regulated promoter (e.g., transfected with a reporter gene construct) can be assayed in an HTS format to identify DNA-binding proteins that regulate the promoter. In the assay's simplest form, inhibition of a negative regulator results in an increase of the signal and inhibition of a positive regulator results in a decrease of the signal.

If a characteristic of an organism is determined to be genetically linked to a polymorphism through RFLP or QTL analysis, the present invention can be used to gain insight regarding whether that genetic polymorphism might be directly responsible for the

characteristic. For example, a fragment defining the genetic polymorphism or sequences in the vicinity of such a genetic polymorphism can be amplified to produce an RNA, the duplex RNA can be introduced to the organism, and whether an alteration in the characteristic is correlated with inhibition can be determined. Of course, there may be trivial explanations for negative results with this type of assay, for example: inhibition of the target gene causes lethality, inhibition of the target gene may not result in any observable alteration, the fragment contains nucleotide sequences that are not capable of inhibiting the target gene, or the target gene's activity is redundant.

The present invention may be useful in allowing the inhibition of essential genes. Such genes may be required for cell or organism viability at only particular stages of development or cellular compartments. The functional equivalent of conditional mutations may be produced by inhibiting activity of the target gene when or where it is not required for viability. The invention allows addition of RNA at specific times of development and locations in the organism without introducing permanent mutations into the target genome.

If alternative splicing produced a family of transcripts that were distinguished by usage of characteristic exons, the present invention can target inhibition through the appropriate exons to specifically inhibit or to distinguish among the functions of family members. For example, a hormone that contained an alternatively spliced transmembrane domain may be expressed in both membrane bound and secreted forms. Instead of isolating a nonsense mutation that terminates translation before the transmembrane domain, the functional consequences of having only secreted hormone can be determined according to the invention by targeting the exon containing the transmembrane domain and thereby inhibiting expression of membrane-bound hormone.

The present invention may be used alone or as a component of a kit having at least one of the reagents necessary to carry out the *in vitro* or *in vivo* introduction of RNA to test samples or subjects. Preferred components are the dsRNA and a vehicle that promotes introduction of the dsRNA. Such a kit may also include instructions to allow a user of the kit to practice the invention.

Pesticides may include the RNA molecule itself, an expression construct capable of expressing the RNA, or organisms transfected with the expression construct. The

pesticide of the present invention may serve as an arachnicide, insecticide, nematicide, viricide, bactericide, and/or fungicide. For example, plant parts that are accessible above ground (e.g., flowers, fruits, buds, leaves, seeds, shoots, bark, stems) may be sprayed with pesticide, the soil may be soaked with pesticide to access plant parts growing beneath ground level, or the pest may be contacted with pesticide directly. If pests interact with each other, the RNA may be transmitted between them. Alternatively, if inhibition of the target gene results in a beneficial effect on plant growth or development, the aforementioned RNA, expression construct, or transfected organism may be considered a nutritional agent. In either case, genetic engineering of the plant is not required to achieve the objectives of the invention.

Alternatively, an organism may be engineered to produce dsRNA which produces commercially or medically beneficial results, for example, resistance to a pathogen or its pathogenic effects, improved growth, or novel developmental patterns.

Used as either an pesticide or nutrient, a formulation of the present invention may be delivered to the end user in dry or liquid form: for example, as a dust, granulate, emulsion, paste, solution, concentrate, suspension, or encapsulation. Instructions for safe and effective use may also be provided with the formulation. The formulation might be used directly, but concentrates would require dilution by mixing with an extender provided by the formulator or the end user. Similarly, an emulsion, paste, or suspension may require the end user to perform certain preparation steps before application. The formulation may include a combination of chemical additives known in the art such as solid carriers, minerals, solvents, dispersants, surfactants, emulsifiers, tackifiers, binders, and other adjuvants. Preservatives and stabilizers may also be added to the formulation to facilitate storage. The crop area or plant may also be treated simultaneously or separately with other pesticides or fertilizers. Methods of application include dusting, scattering or pouring, soaking, spraying, atomizing, and coating. The precise physical form and chemical composition of the formulation, and its method of application, would be chosen to promote the objectives of the invention and in accordance with prevailing circumstances. Expression constructs and transfected hosts capable of replication may also promote the persistence and/or spread of the formulation.

Description of the dsRNA Inhibition Phenomenon in *C. elegans*

The operation of the present invention was shown in the model genetic organism *Caenorhabditis elegans*.

Introduction of RNA into cells had been seen in certain biological systems to interfere with function of an endogenous gene[1,2]. Many such effects were believed to result from a simple antisense mechanism dependent on hybridization between injected single-stranded RNA and endogenous transcripts. In other cases, a more complex mechanism had been suggested. One instance of an RNA-mediated mechanism was RNA interference (RNAi) phenomenon in the nematode *C. elegans*. RNAi had been used in a variety of studies to manipulate gene expression[3,4].

Despite the usefulness of RNAi in *C. elegans*, many features had been difficult to explain. Also, the lack of a clear understanding of the critical requirements for interfering RNA led to a sporadic record of failure and partial success in attempts to extend RNAi beyond the earliest stages following injection. A statement frequently made in the literature was that sense and antisense RNA preparations are each sufficient to cause interference[3,4]. The only precedent for such a situation was in plants where the process of co-suppression had a similar history of usefulness in certain cases, failure in others, and no ability to design interference protocols with a high chance of success. Working with *C. elegans*, we discovered an RNA structure that would give effective and uniform genetic inhibition. The prior art did not teach or suggest that RNA structure was a critical feature for inhibition of gene expression. Indeed the ability of crude sense and antisense preparations to produce interference[3,4] had been taken as an indication that RNA structure was not a critical factor. Instead, the extensive plant literature and much of the ongoing research in *C. elegans* was focused on the possibility that detailed features of the target gene sequence or its chromosomal locale was the critical feature for interfering with gene expression.

The inventors carefully purified sense or antisense RNA for *unc-22* and tested each for gene-specific inhibition. While the crude sense and antisense preparations had strong interfering activity, it was found that the purified sense and antisense RNAs had only marginal inhibitory activity. This was unexpected because many techniques in molecular biology are based on the assumption that RNA produced with specific *in vitro*

promoters (e.g., T3 or T7 RNA polymerase), or with characterized promoters *in vivo*, is produced predominantly from a single strand. The inventors had carried out purification of these crude preparations to investigate whether a small fraction of the RNA had an unusual structure which might be responsible for the observed genetic inhibition. To rigorously test whether double-stranded character might contribute to genetic inhibition, the inventors carried out additional purification of single-stranded RNAs and compared inhibitory activities of individual strands with that of the double-stranded hybrid.

The following examples are meant to be illustrative of the present invention; however, the practice of the invention is not limited or restricted in any way by them.

Analysis of RNA-Mediated Inhibition of *C. elegans* Genes

The *unc-22* gene was chosen for initial comparisons of activity as a result of previous genetic analysis that yields a semi-quantitative comparison between *unc-22* gene activity and the movement phenotypes of animals[3,8]: decreases in activity produce an increasingly severe twitching phenotype, while complete loss of function results in the additional appearance of muscle structural defects and impaired motility. *unc-22* encodes an abundant but non-essential myofilament protein[7-9]. *unc-22* mRNA is present at several thousand copies per striated muscle cell[3].

Purified antisense and sense RNAs covering a 742 nt segment of *unc-22* had only marginal inhibitory activity, requiring a very high dose of injected RNA for any observable effect (Figure 4). By contrast, a sense+antisense mixture produced a highly effective inhibition of endogenous gene activity (Figure 4). The mixture was at least two orders of magnitude more effective than either single strand in inhibiting gene expression. The lowest dose of the sense+antisense mixture tested, approximately 60,000 molecules of each strand per adult, led to twitching phenotypes in an average of 100 progeny. *unc-22* expression begins in embryos with approximately 500 cells. At this point, the original injected material would be diluted to at most a few molecules per cell.

The potent inhibitory activity of the sense+antisense mixture could reflect formation of double-stranded RNA (dsRNA), or conceivably some alternate synergy between the strands. Electrophoretic analysis indicated that the injected material was predomi-

nantly double stranded. The dsRNA was gel purified from the annealed mixture and found to retain potent inhibitory activity. Although annealing prior to injection was compatible with inhibition, it was not necessary. Mixing of sense and antisense RNAs in low salt (under conditions of minimal dsRNA formation), or rapid sequential injection of sense and antisense strands, were sufficient to allow complete inhibition. A long interval (>1 hour) between sequential injections of sense and antisense RNA resulted in a dramatic decrease in inhibitory activity. This suggests that injected single strands may be degraded or otherwise rendered inaccessible in the absence of the complementary strand.

An issue of specificity arises when considering known cellular responses to dsRNA. Some organisms have a dsRNA-dependent protein kinase that activates a panic response mechanism[10]. Conceivably, the inventive sense+antisense synergy could reflect a non-specific potentiation of antisense effects by such a panic mechanism. This was not found to be the case: co-injection of dsRNA segments unrelated to *unc-22* did not potentiate the ability of *unc-22* single strands to mediate inhibition. Also investigated was whether double-stranded structure could potentiate inhibitory activity when placed in *cis* to a single-stranded segment. No such potentiation was seen; unrelated double-stranded sequences located 5' or 3' of a single-stranded *unc-22* segment did not stimulate inhibition. Thus potentiation of gene-specific inhibition was observed only when dsRNA sequences exist within the region of homology with the target gene.

The phenotype produced by *unc-22* dsRNA was specific. Progeny of injected animals exhibited behavior indistinguishable from characteristic *unc-22* loss of function mutants. Target-specificity of dsRNA effects using three additional genes with well characterized phenotypes (Figure 1 and Table 1). *unc-54* encodes a body wall muscle myosin heavy chain isoform required for full muscle contraction[7,11,12], *fem-1* encodes an ankyrin-repeat containing protein required in hermaphrodites for sperm production[13,14], and *hlh-1* encodes a *C. elegans* homolog of the myoD family required for proper body shape and motility[15,16]. For each of these genes, injection of dsRNA produced progeny broods exhibiting the known null mutant phenotype, while the purified single strands produced no significant reduction in gene expression. With one exception, all of the phenotypic consequences of dsRNA injection were those expected from inhibition of the corresponding gene. The exception (segment *unc54C*, which led to an embryonic and

larval arrest phenotype not seen with *unc-54* null mutants) was illustrative. This segment covers the highly conserved myosin motor domain, and might have been expected to inhibit the activity of other highly related myosin heavy chain genes[17]. This interpretation would support uses of the present invention in which nucleotide sequence comparison of dsRNA and target gene show less than 100% identity. The *unc54C* segment has been unique in our overall experience to date: effects of 18 other dsRNA segments have all been limited to those expected from characterized null mutants.

The strong phenotypes seen following dsRNA injection are indicative of inhibitory effects occurring in a high fraction of cells. The *unc-54* and *hlh-1* muscle phenotypes, in particular, are known to result from a large number of defective muscle cells[11,16]. To examine inhibitory effects of dsRNA on a cellular level, a transgenic line expressing two different GFP-derived fluorescent reporter proteins in body muscle was used. Injection of dsRNA directed to *gfp* produced dramatic decreases in the fraction of fluorescent cells (Figure 2). Both reporter proteins were absent from the negative cells, while the few positive cells generally expressed both GFP forms.

The pattern of mosaicism observed with *gfp* inhibition was not random. At low doses of dsRNA, the inventors saw frequent inhibition in the embryonically-derived muscle cells present when the animal hatched. The inhibitory effect in these differentiated cells persisted through larval growth: these cells produced little or no additional GFP as the affected animals grew. The 14 postembryonically-derived striated muscles are born during early larval stages and were more resistant to inhibition. These cells have come through additional divisions (13-14 versus 8-9 for embryonic muscles[18,19]). At high concentrations of *gfp* dsRNA, inhibition was noted in virtually all striated bodywall muscles, with occasional single escaping cells including cells born in embryonic or post-embryonic stages. The nonstriated vulval muscles, born during late larval development, appeared resistant to genetic inhibition at all tested concentrations of injected RNA. The latter result is important for evaluating the use of the present invention in other systems. First, it indicates that failure in one set of cells from an organism does not necessarily indicate complete non-applicability of the invention to that organism. Second, it is important to realize that not all tissues in the organism need to be affected for the invention to be used in an organism. This may serve as an advantage in some situations.

A few observations serve to clarify the nature of possible targets and mechanisms for RNA-mediated genetic inhibition in *C. elegans*:

First, dsRNA segments corresponding to a variety of intron and promoter sequences did not produce detectable inhibition (Table 1). Although consistent with possible inhibition at a post-transcriptional level, these experiments do not rule out inhibition at the level of the gene.

Second, dsRNA injection produced a dramatic decrease in the level of the endogenous mRNA transcript (Figure 3). Here, a *mex-3* transcript that is abundant in the gonad and early embryos[20] was targeted, where straightforward *in situ* hybridization can be performed[5]. No endogenous *mex-3* mRNA was observed in animals injected with a dsRNA segment derived from *mex-3* (Figure 3D), but injection of purified *mex-3* antisense RNA resulted in animals that retained substantial endogenous mRNA levels (Figure 3C).

Third, dsRNA-mediated inhibition showed a surprising ability to cross cellular boundaries. Injection of dsRNA for *unc-22*, *gfp*, or *lacZ* into the body cavity of the head or tail produced a specific and robust inhibition of gene expression in the progeny brood (Table 2). Inhibition was seen in the progeny of both gonad arms, ruling out a transient "nicking" of the gonad in these injections. dsRNA injected into body cavity or gonad of young adults also produced gene-specific inhibition in somatic tissues of the injected animal (Table 2).

Table 3 shows that *C. elegans* can respond in a gene-specific manner to dsRNA encountered in the environment. Bacteria are a natural food source for *C. elegans*. The bacteria are ingested, ground in the animal's pharynx, and the bacterial contents taken up in the gut. The results show that *E. coli* bacteria expressing dsRNAs can confer specific inhibitory effects on *C. elegans* nematode larvae that feed on them.

Three *C. elegans* genes were analyzed. For each gene, corresponding dsRNA was expressed in *E. coli* by inserting a segment of the coding region into a plasmid construct designed for bidirectional transcription by bacteriophage T7 RNA polymerase. The dsRNA segments used for these experiments were the same as those used in previous microinjection experiments (see Figure 1). The effects resulting from feeding these bacteria to *C. elegans* were compared to the effects achieved by microinjecting animals

with dsRNA.

The *C. elegans* gene *unc-22* encodes an abundant muscle filament protein. *unc-22* null mutations produce a characteristic and uniform twitching phenotype in which the animals can sustain only transient muscle contraction. When wild-type animals were fed bacteria expressing a dsRNA segment from *unc-22*, a high fraction (85%) exhibited a weak but still distinct twitching phenotype characteristic of partial loss of function for the *unc-22* gene. The *C. elegans fem-1* gene encodes a late component of the sex determination pathway. Null mutations prevent the production of sperm and lead euploid (XX) animals to develop as females, while wild type XX animals develop as hermaphrodites. When wild-type animals were fed bacteria expressing dsRNA corresponding to *fem-1*, a fraction (43%) exhibit a sperm-less (female) phenotype and were sterile. Finally, the ability to inhibit gene expression of a transgene target was assessed. When animals carrying a *gfp* transgene were fed bacteria expressing dsRNA corresponding to the *gfp* reporter, an obvious decrease in the overall level of GFP fluorescence was observed, again in approximately 12% of the population (see Figure 5, panels B and C).

The effects of these ingested RNAs were specific. Bacteria carrying different dsRNAs from *fem-1* and *gfp* produced no twitching, dsRNAs from *unc-22* and *fem-1* did not reduce *gfp* expression, and dsRNAs from *gfp* and *unc-22* did not produce females. These inhibitory effects were apparently mediated by dsRNA: bacteria expressing only the sense or antisense strand for either *gfp* or *unc-22* caused no evident phenotypic effects on their *C. elegans* predators.

Table 4 shows the effects of bathing *C. elegans* in a solution containing dsRNA. Larvae were bathed for 24 hours in solutions of the indicated dsRNAs (1 mg/ml), then allowed to recover in normal media and allowed to grow under standard conditions for two days. The *unc-22* dsRNA was segment ds-*unc22A* from Figure 1. *pos*-1 and *sqt*-3 dsRNAs were from the full length cDNA clones. *pos-1* encodes an essential maternally provided component required early in embyogenesis. Mutations removing *pos-1* activity have an early embryonic arrest characteristic of *skn*-like mutations[29, 30]. Cloning and activity patterns for *sqt-3* have been described[31]. *C. elegans sqt-3* mutants have mutations in the *col-1* collagen gene[31]. Phenotypes of affected animals are noted. Incidences of

clear phenotypic effects in these experiments were 5-10% for *unc-22*, 50% for *pos-1*, and 5% for *sqt-3*. These are frequencies of unambiguous phenocopies; other treated animals may have had marginal defects corresponding to the target gene that were not observable. Each treatment was fully gene-specific in that *unc-22* dsRNA produced only Unc-22 phenotypes, *pos-1* dsRNA produced only Pos-1 phenotypes, and *sqt-3* dsRNA produced only Sqt-3 phenotypes.

Some of the results described herein were published after the filing of our provisional application. Those publications and a review can be cited as Fire, A., et al. Nature, 391, 806-811, 1998; Timmons, L. & Fire, A. Nature, 395, 854, 1998; and Montgomery, M.K. & Fire, A. Trends in Genetics, 14, 255-258, 1998.

The effects described herein significantly augment available tools for studying gene function in *C. elegans* and other organisms. In particular, functional analysis should now be possible for a large number of interesting coding regions[21] for which no specific function have been defined. Several of these observations show the properties of dsRNA that may affect the design of processes for inhibition of gene expression. For example, one case was observed in which a nucleotide sequence shared between several myosin genes may inhibit gene expression of several members of a related gene family.

Methods of RNA Synthesis and Microinjection

RNA was synthesized from phagemid clones with T3 and T7 RNA polymerase[6], followed by template removal with two sequential DNase treatments. In cases where sense, antisense, and mixed RNA populations were to be compared, RNAs were further purified by electrophoresis on low-gelling-temperature agarose. Gel-purified products appeared to lack many of the minor bands seen in the original "sense" and "antisense" preparations. Nonetheless, RNA species accounting for less than 10% of purified RNA preparations would not have been observed. Without gel purification, the "sense" and "antisense" preparations produced significant inhibition. This inhibitory activity was reduced or eliminated upon gel purification. By contrast, sense+antisense mixtures of gel purified and non-gel-purified RNA preparations produced identical effects.

Following a short (5 minute) treatment at 68°C to remove secondary structure, sense+antisense annealing was carried out in injection buffer[27] at 37°C for 10-30 minutes.

Formation of predominantly double stranded material was confirmed by testing migration on a standard (non-denaturing) agarose gel: for each RNA pair, gel mobility was shifted to that expected for double-stranded RNA of the appropriate length. Co-incubation of the two strands in a low-salt buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA) was insufficient for visible formation of double-stranded RNA *in vitro*. Non-annealed sense+antisense RNAs for *unc22B* and *gfpG* were tested for inhibitory effect and found to be much more active than the individual single strands, but 2-4 fold less active than equivalent pre-annealed preparations.

After pre-annealing of the single strands for *unc22A*, the single electrophoretic species corresponding in size to that expected for dsRNA was purified using two rounds of gel electrophoresis. This material retained a high degree of inhibitory activity.

Except where noted, injection mixes were constructed so animals would receive an average of $0.5x10^6$ to $1.0x10^6$ molecules of RNA. For comparisons of sense, antisense, and dsRNA activities, injections were compared with equal masses of RNA (i.e., dsRNA at half the molar concentration of the single strands). Numbers of molecules injected per adult are given as rough approximations based on concentration of RNA in the injected material (estimated from ethidium bromide staining) and injection volume (estimated from visible displacement at the site of injection). A variability of several-fold in injection volume between individual animals is possible; however, such variability would not affect any of the conclusions drawn herein.

Methods for Analysis of Phenotypes

Inhibition of endogenous genes was generally assayed in a wild type genetic background (N2). Features analyzed included movement, feeding, hatching, body shape, sexual identity, and fertility. Inhibition with *gfp* [27] and *lacZ* activity was assessed using strain PD4251. This strain is a stable transgenic strain containing an integrated array (ccIs4251) made up of three plasmids: pSAK4 (*myo-3* promoter driving mitochondrially targeted GFP), pSAK2 (*myo-3* promoter driving a nuclear targeted GFP-LacZ fusion), and a *dpy-20* subclone[26] as a selectable marker. This strain produces GFP in all body muscles, with a combination of mitochondrial and nuclear localization. The two distinct compartments are easily distinguished in these cells, allowing a facile distinction between

cells expressing both, either, or neither of the original GFP constructs.

Gonadal injection was performed by inserting the microinjection needle into the gonadal syncitium of adults and expelling 20-100 pl of solution (see Reference 25). Body cavity injections followed a similar procedure, with needle insertion into regions of the head and tail beyond the positions of the two gonad arms. Injection into the cytoplasm of intestinal cells was another effective means of RNA delivery, and may be the least disruptive to the animal. After recovery and transfer to standard solid media, injected animals were transferred to fresh culture plates at 16 hour intervals. This yields a series of semi-synchronous cohorts in which it was straightforward to identify phenotypic differences. A characteristic temporal pattern of phenotypic severity is observed among progeny. First, there is a short "clearance" interval in which unaffected progeny are produced. These include impermeable fertilized eggs present at the time of injection. After the clearance period, individuals are produced which show the inhibitory phenotype. After injected animals have produced eggs for several days, gonads can in some cases "revert" to produce incompletely affected or phenotypically normal progeny.

Additional Description of the Results

Figure 1 shows genes used to study RNA-mediated genetic inhibition in *C. elegans*. Intron-exon structure for genes used to test RNA-mediated inhibition are shown (exons: filled boxes; introns: open boxes; 5' and 3' untranslated regions: shaded; sequence references are as follows: *unc-22* [9], *unc-54* [12], *fem-1* [14], and *hlh-1* [15]). These genes were chosen based on: (1) a defined molecular structure, (2) classical genetic data showing the nature of the null phenotype. Each segment tested for inhibitory effects is designated with the name of the gene followed by a single letter (e.g., *unc22C*). Segments derived from genomic DNA are shown above the gene, segments derived from cDNA are shown below the gene. The consequences of injecting double-stranded RNA segments for each of these genes is described in Table 1. dsRNA sequences from the coding region of each gene produced a phenotype resembling the null phenotype for that gene.

The effects of inhibitory RNA were analyzed in individual cells (Figure 2, panels A-H). These experiments were carried out in a reporter strain (called PD4251) expressing

two different reporter proteins: nuclear GFP-LacZ and mitochondrial GFP, both expressed in body muscle. The fluorescent nature of these reporter proteins allowed us to examine individual cells under the fluorescence microscope to determine the extent and generality of the observed inhibition of gene. ds-*unc22A* RNA was injected as a negative control. GFP expression in progeny of these injected animals was not affected. The GFP patterns of these progeny appeared identical to the parent strain, with prominent fluorescence in nuclei (the nuclear localized GFP-LacZ) and mitochondria (the mitochondrially targeted GFP): young larva (Figure 2A), adult (Figure 2B), and adult body wall at high magnification (Figure 2C).

In contrast, the progeny of animals injected with ds-*gfpG* RNA are affected (Figures 2D-F). Observable GFP fluorescence is completely absent in over 95% of the cells. Few active cells were seen in larvae (Figure 2D shows a larva with one active cell; uninjected controls show GFP activity in all 81 body wall muscle cells). Inhibition was not effective in all tissues: the entire vulval musculature expressed active GFP in an adult animal (Figure 2E). Rare GFP positive body wall muscle cells were also seen adult animals (two active cells are shown in Figure 2F). Inhibition was target specific (Figures 2G-I). Animals were injected with ds-*lacZL* RNA, which should affect the nuclear but not the mitochondrial reporter construct. In the animals derived from this injection, mitochondrial-targeted GFP appeared unaffected while the nuclear-targeted GFP-LacZ was absent from almost all cells (larva in Figure 2G). A typical adult lacked nuclear GFP-LacZ in almost all body-wall muscles but retained activity in vulval muscles (Figure 2H). Scale bars in Figure 2 are 20 μm.

The effects of double-stranded RNA corresponding to *mex-3* on levels of the endogenous mRNA was shown by *in situ* hybridization to embryos (Figure 3, panels A-D). The 1262 nt *mex-3* cDNA clone[20] was divided into two segments, *mex-3*A and *mex-3*B with a short (325 nt) overlap. Similar results were obtained in experiments with no overlap between inhibiting and probe segments. *mex-3*B antisense or dsRNA was injected into the gonads of adult animals, which were maintained under standard culture conditions for 24 hours before fixation and *in situ* hybridization (see Reference 5). The *mex-3*B dsRNA produced 100% embryonic arrest, while >90% of embryos from the antisense injections hatched. Antisense probes corresponding to *mex-3*A were used to

assay distribution of the endogenous *mex-3* mRNA (dark stain). Four-cell stage embryos were assayed; similar results were observed from the 1 to 8 cell stage and in the germline of injected adults. The negative control (the absence of hybridization probe) showed a lack of staining (Figure 3A). Embryos from uninjected parents showed a normal pattern of endogenous *mex-3* RNA (Figure 3B). The observed pattern of *mex-3* RNA was as previously described in Reference 20. Injection of purified *mex-3*B antisense RNA produced at most a modest effect: the resulting embryos retained *mex-3* mRNA, although levels may have been somewhat less than wild type (Figure 3C). In contrast, no *mex-3* RNA was detected in embryos from parents injected with dsRNA corresponding to *mex-3*B (Figure 3D). The scale of Figure 3 is such that each embryo is approximately 50 μm in length.

Gene-specific inhibitory activity by *unc-22A* RNA was measured as a function of RNA structure and concentration (Figure 4). Purified antisense and sense RNA from *unc22A* were injected individually or as an annealed mixture. "Control" was an unrelated dsRNA (*gfpG*). Injected animals were transferred to fresh culture plates 6 hours (columns labeled **1**), 15 hours (columns labeled **2**), 27 hours (columns labeled **3**), 41 hours (columns labeled **4**), and 56 hours (columns labeled **5**) after injection. Progeny grown to adulthood were scored for movement in their growth environment, then examined in 0.5 mM levamisole. The main graph indicates fractions in each behavioral class. Embryos in the uterus and already covered with an eggshell at the time of injection were not affected and, thus, are not included in the graph. The bottom-left diagram shows the genetically derived relationship between *unc-22* gene dosage and behavior based on analyses of *unc-22* heterozygotes and polyploids[8,3].

Figures 5 A-C show a process and examples of genetic inhibition following ingestion by *C. elegans* of dsRNAs from expressing bacteria. A general strategy for production of dsRNA is to clone segments of interest between flanking copies of the bacteriophage T7 promoter into a bacterial plasmid construct (Figure 5A). A bacterial strain (BL21/DE3)[28] expressing the T7 polymerase gene from an inducible (Lac) promoter was used as a host. A nuclease-resistant dsRNA was detected in lysates of transfected

bacteria. Comparable inhibition results were obtained with the two bacterial expression systems. A GFP-expressing *C. elegans* strain, PD4251 (see Figure 2), was fed on a native bacterial host. These animals show a uniformly high level of GFP fluorescence in body muscles (Figure 5B). PD4251 animals were also reared on a diet of bacteria expressing dsRNA corresponding to the coding region for *gfp*. Under the conditions of this experiment, 12% of these animals showed dramatic decreases in GFP (Figure 5C). As an alternative strategy, single copies of the T7 promoter were used to drive expression of an inverted-duplication for a segment of the target gene, either *unc-22* or *gfp*. This was comparably effective.

All references (e.g., books, articles, applications, and patents) cited in this specification are indicative of the level of skill in the art and their disclosures are incorporated herein in their entirety.

1. Izant, J. & Weintraub, H. *Cell* 36, 1007-1015 (1984).
2. Nellen, W. & Lichtenstein, C. *TIBS* 18, 419-423 (1993).
3. Fire, A., et al. *Development* 113, 503-514 (1991).
4. Guo, S. & Kemphues, K. *Cell* 81, 611-620 (1995).
5. Seydoux, G. & Fire, A. *Development* 120, 2823-2834 (1994).
6. Ausubel, F., et al. *Current Protocols in Molecular Biology*, John Wiley N.Y. (1990).
7. Brenner, S. *Genetics* 77, 71-94 (1974).
8. Moerman, D. & Baillie, D. *Genetics* 91, 95-104 (1979).
9. Benian, G., et al. *Genetics* 134, 1097-1104 (1993).
10. Proud, C. *TIBS* 20, 241-246 (1995).
11. Epstein H., et al. *J. Mol. Biol.* 90 291-300 (1974).
12. Kam, J., et al. *Proc. Natl. Acad. Sci. (U.S.A.)* 80, 4253-4257 (1983).
13. Doniach, T. & Hodgkin J.A. *Dev. Biol.* 106, 223-235 (1984).
14. Spence, A., et al. *Cell* 60, 981-990 (1990).
15. Krause, M., et al. *Cell* 63, 907-919 (1990).
16. Chen, L., et al. *Development*, 120, 1631-1641 (1994).
17. Dibb, N. J., et al. *J. Mol. Biol.* 205, 603-613 (1989).

18. Sulston, J., et al. *Dev. Biol.* 100, 64-119 (1983).

19. Sulston, J. & Horvitz, H. *Dev. Biol.* 82, 41-55 (1977).

20. Draper B.W., et al. *Cell* 87, 205-216 (1996).

21. Sulston, J., et al. *Nature* 356, 37-41 (1992).

22. Matzke, M. & Matzke, A. *Plant Physiol.* 107, 679-685 (1995).

23. Ratcliff, F., et al. *Science* 276, 1558-1560 (1997).

24. Latham, K. *Trends in Genetics* 12,134-138 (1996).

25. Mello, C. & Fire, A. *Methods in Cell Biology* 48, 451-482 (1995).

26. Clark, D., et al. *Mol. Gen. Genet.* 247, 367-378 (1995).

27. Chalfie, M., et al. *Science* 263, 802-805 (1994).

28. Studier, F., et al. *Methods in Enzymology* 185, 60-89 (1990).

29. Bowerman, B., et al. *Cell* 68, 1061-1075 (1992).

30. Mello, C.C., et al. *Cell* 70, 163-176 (1992).

31. van der Keyl, H., et al. *Develop. Dynamics* 201, 86-94 (1994).

32. Goeddel, D.V. *Gene Expression Technology*, Academic Press, 1990.

33. Kriegler, M. *Gene Transfer and Expression*, Stockton Press, 1990.

34. Murray, E.J. *Gene Transfer and Expression Protocols*, Humana Press, 1991.

Table 1. Effects of sense, antisense, and mixed RNAs on progeny of injected animals.

Gene and Segment		Size	Injected RNA	F1 Phenotype
unc-22			***unc-22* null mutants: strong twitchers**[7,8]	
unc22A[a]	exon 21-22	742	sense	wild type
			antisense	wild type
			sense+antisense	strong twitchers (100%)
unc22B	exon 27	1033	sense	wild type
			antisense	wild type
			sense+antisense	strong twitchers (100%)
unc22C	exon 21-22[b]	785	sense+antisense	strong twitchers (100%)
fem-1			***fem-1* null mutants: female (no sperm)**[13]	
fem1A	exon 10[c]	531	sense	hermaphrodite (98%)
			antisense	hermaphrodite (>98%)
			sense+antisense	female (72%)
fem1B	intron 8	556	sense+antisense	hermaphrodite (>98%)
unc-54			***unc-54* null mutants: paralyzed**[7,11]	
unc54A	exon 6	576	sense	wild type (100%)
			antisense	wild type (100%)
			sense+antisense	paralyzed (100%)[d]
unc54B	exon 6	651	sense	wild type (100%)
			antisense	wild type (100%)
			sense+antisense	paralyzed (100%)[d]
unc54C	exon 1-5	1015	sense+antisense	arrested embryos and larvae (100%)
unc54D	promoter	567	sense+antisense	wild type (100%)
unc54E	intron 1	369	sense+antisense	wild type (100%)
unc54F	intron 3	386	sense+antisense	wild type (100%)

Table 1 (continued).

Gene and Segment		Size	Injected RNA	F1 Phenotype
hlh-1			***hlh-1* null mutants: lumpy-dumpy larvae**[16]	
hlh1A	exons 1-6	1033	sense	wild type (<2% lpy-dpy)
			antisense	wild type (<2% lpy-dpy)
			sense+antisense	lpy-dpy larvae (>90%)[e]
hlh1B	exons 1-2	438	sense+antisense	lpy-dpy larvae (>80%)[e]
hlh1C	exons 4-6	299	sense+antisense	lpy-dpy larvae (>80%)[e]
hlh1D	intron 1	697	sense+antisense	wild type (<2% lpy-dpy)
myo-3 driven GFP transgenes [f]				
myo-3::NLS::gfp::lacZ			**makes nuclear GFP in body muscle**	
gfpG	exons 2-5	730	sense	nuclear GFP-LacZ pattern of parent strain
			antisense	nuclear GFP-LacZ pattern of parent strain
			sense+antisense	nuclear GFP-LacZ absent in 98% of cells
lacZL	exon 12-14	830	sense+antisense	nuclear GFP-LacZ absent in >95% of cells
myo-3::MtLS::gfp			**makes mitochondrial GFP in body muscle**	
gfpG	exons 2-5	730	sense	mitochondrial GFP pattern of parent strain
			antisense	mitochondrial GFP pattern of parent strain
			sense+antisense	mitochondrial GFP absent in 98% of cells
lacZL	exon 12-14	830	sense+antisense	mitochondrial GFP pattern of parent strain

Legend of Table 1

Each RNA was injected into 6-10 adult hermaphrodites ($0.5\text{-}1 \times 10^6$ molecules into each gonad arm). After 4-6 hours (to clear pre-fertilized eggs from the uterus) injected animals were transferred and eggs collected for 20-22 hours. Progeny phenotypes were scored upon hatching and subsequently at 12-24 hour intervals.

a: To obtain a semi-quantitative assessment of the relationship between RNA dose and phenotypic response, we injected each *unc22A* RNA preparation at a series of different concentrations. At the highest dose tested (3.6×10^6 molecules per gonad), the

individual sense and antisense *unc22A* preparations produced some visible twitching (1% and 11% of progeny respectively). Comparable doses of ds-*unc22A* RNA produced visible twitching in all progeny, while a 120-fold lower dose of ds-*unc22A* RNA produced visible twitching in 30% of progeny.

b: *unc22C* also carries the intervening intron (43 nt).

c: *fem1A* also carries a portion (131 nt) of intron 10.

d: Animals in the first affected broods (laid at 4-24 hours after injection) showed movement defects indistinguishable from those of null mutants in *unc-54*. A variable fraction of these animals (25-75%) failed to lay eggs (another phenotype of *unc-54* null mutants), while the remainder of the paralyzed animals were egg-laying positive. This may indicate partial inhibition of *unc-54* activity in vulval muscles. Animals from later broods frequently exhibit a distinct partial loss-of-function phenotype, with contractility in a subset of body wall muscles.

e: Phenotypes of *hlh-1* inhibitory RNA include arrested embryos and partially elongated L1 larvae (the *hlh-1* null phenotype) seen in virtually all progeny from injection of ds-*hlh1A* and about half of the affected animals from ds-*hlh1B* and ds-*hlh1C*) and a set of less severe defects (seen with the remainder of the animals from ds-*hlh1B* and ds-*hlh1C*). The less severe phenotypes are characteristic of partial loss of function for *hlh-1*.

f: The host for these injections, PD4251, expresses both mitochondrial GFP and nuclear GFP-LacZ. This allows simultaneous assay for inhibition of *gfp* (loss of all fluorescence) and *lacZ* (loss of nuclear fluorescence). The table describes scoring of animals as L1 larvae. ds-*gfpG* caused a loss of GFP in all but 0-3 of the 85 body muscles in these larvae. As these animals mature to adults, GFP activity was seen in 0-5 additional bodywall muscles and in the eight vulval muscles.

Table 2. Effect of injection point on genetic inhibition in injected animals and their progeny.

dsRNA	Site of injection	Injected animal phenotype	Progeny Phenotype
None	gonad or body cavity	no twitching	no twitching
None	gonad or body cavity	strong nuclear & mitochondrial GFP	strong nuclear & mitochondrial GFP
unc22B	Gonad	weak twitchers	strong twitchers
unc22B	Body Cavity Head	weak twitchers	strong twitchers
unc22B	Body Cavity Tail	weak twitchers	strong twitchers
gfpG	Gonad	lower nuclear & mitochondrial GFP	rare or absent nuclear & mitochondrial GFP
gfpG	Body Cavity Tail	lower nuclear & mitochondrial GFP	rare or absent nuclear & mitochondrial GFP
lacZL	Gonad	lower nuclear GFP	rare or absent nuclear GFP
lacZL	Body Cavity Tail	lower nuclear GFP	rare or absent nuclear GFP

Table 3. *C. elegans* can respond in a gene-specific manner to environmental dsRNA.

Bacterial Food	Movement	Germline Phenotype	GFP-Transgene Expression
BL21(DE3)	0% twitch	< 1% female	< 1% faint GFP
BL21(DE3) [*fem-1* dsRNA]	0% twitch	43% female	< 1% faint GFP
BL21(DE3) [*unc22* dsRNA]	85% twitch	< 1% female	< 1% faint GFP
BL21(DE3) [*gfp* dsRNA]	0% twitch	< 1% female	12% faint GFP

Table 4. Effects of bathing *C. elegans* in a solution containing dsRNA.

dsRNA	Biological Effect
unc-22	Twitching (similar to partial loss of *unc-22* function)
pos-1	Embryonic arrest (similar to loss of *pos-1* function)
sqt-3	Shortened body (Dpy) (similar to partial loss of *sqt-3* function)

In Table 2, gonad injections were carried out into the GFP reporter strain PD4251, which expresses both mitochondrial GFP and nuclear GFP-LacZ. This allowed simultaneous assay of inhibition with *gfp* (fainter overall fluorescence), *lacZ* (loss of nuclear fluorescence), and *unc-22* (twitching). Body cavity injections were carried out into the tail region, to minimize accidental injection of the gonad; equivalent results have been observed with injections into the anterior region of the body cavity. An equivalent set of injections was also performed into a single gonad arm. For all sites of injection, the entire progeny brood showed phenotypes identical to those described in Table 1. This included progeny produced from both injected and uninjected gonad arms. Injected animals were scored three days after recovery and showed somewhat less dramatic phenotypes than their progeny. This could in part be due to the persistence of products already present in the injected adult. After ds-*unc22B* injection, a fraction of the injected animals twitch weakly under standard growth conditions (10 out of 21 animals). Levamisole treatment led to twitching of 100% (21/21) of these animals. Similar effects were seen with ds-*unc22A*. Injections of ds-*gfpG* or ds-*lacZL* produced a dramatic decrease (but not elimination) of the corresponding GFP reporters. In some cases, isolated cells or parts of animals retained strong GFP activity. These were most frequently seen in the anterior region and around the vulva. Injections of ds-*gfpG* and ds-*lacZL* produced no twitching, while injections of ds-*unc22A* produced no change in GFP fluorescence pattern.

While the present invention has been described in connection with what is presently considered to be practical and preferred embodiments, it is understood that the invention is not to be limited or restricted to the disclosed embodiments but, on the contrary, is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

Thus it is to be understood that variations in the described invention will be obvious to those skilled in the art without departing from the novel aspects of the present invention and such variations are intended to come within the scope of the present invention.

WE CLAIM:

1. A method to inhibit expression of a target gene in a cell comprising introduction of a ribonucleic acid (RNA) into the cell in an amount sufficient to inhibit expression of the target gene, wherein the RNA comprises a double-stranded structure with an identical nucleotide sequence as compared to a portion of the target gene.

2. The method of claim 1 in which the target gene is a cellular gene.

3. The method of claim 1 in which the target gene is an endogenous gene.

4. The method of claim 1 in which the target gene is a transgene.

5. The method of claim 1 in which the target gene is a viral gene.

6. The method of claim 1 in which the cell is from an animal.

7. The method of claim 1 in which the cell is from a plant.

8. The method of claim 6 in which the cell is from an invertebrate animal.

9. The method of claim 8 in which the cell is from a nematode.

10. The method of claim 1 in which the identical nucleotide sequence is at least 50 bases in length.

11. The method of claim 1 in which the target gene expression is inhibited by at least 10%.

12. The method of claim 1 in which the cell is present in an organism and inhibition of target gene expression demonstrates a loss-of function phenotype.

13. The method of claim 1 in which the RNA comprises one strand which is self-complementary .

14. The method of claim 1 in which the RNA comprises two separate complementary strands.

15. The method of claim 14 further comprising synthesis of the two complementary strands and initiation of RNA duplex formation outside the cell.

16. The method of claim 14 further comprising synthesis of the two complementary strands and initiation of RNA duplex formation inside the cell.

17. The method of claim 1 in which the cell is present in an organism, and the RNA is introduced within a body cavity of the organism and outside the cell.

18. The method of claim 1 in which the cell is present in an organism and the RNA is introduced by extracellular injection into the organism.

19. The method of claim 1 in which the cell is present in a first organism, and the RNA is introduced to the first organism by feeding a second, RNA-containing organism to the first organism.

20. The method of claim 19 in which the second organism is engineered to produce an RNA duplex.

21. The method of claim 1 in which an expression construct in the cell produces the RNA.

22. A method to inhibit expression of a target gene comprising:

(a) providing an organism containing a target cell, wherein the target cell contains the target gene and the target gene is expressed in the target cell;

(b) contacting a ribonucleic acid (RNA) with the organism, wherein the RNA is comprised of a double-stranded structure with duplexed ribonucleic acid strands and one of the strands is able to duplex with a portion of the target gene; and

(c) introducing the RNA into the target cell, thereby inhibiting expression of the target gene.

23. The method of claim 22 in which the organism is an animal.

24. The method of claim 22 in which the organism is a plant.

25. The method of claim 22 in which the organism is an invertebrate animal.

26. The method of claim 22 in which the organism is a nematode.

27. The method of claim 26 in which a formulation comprised of the RNA is applied on or adjacent to a plant, and disease associated with nematode infection of the plant is thereby reduced.

28. The method of claim 22 in which the identical nucleotide sequence is at least 50 nucleotides in length.

29. The method of claim 22 in which the expression of the target gene is inhibited by at least 10%.

30. The method of claim 22 in which the RNA is introduced within a body cavity of the organism and outside the target cell.

31. The method of claim 22 in which the RNA is introduced by extracellular injection into the organism.

32. The method of claim 22 in which the organism is contacted with the RNA by feeding the organism food containing the RNA.

33. The method of claim 32 in which a genetically-engineered host transcribing the RNA comprises the food.

34. The method of claim 22 in which at least one strand of the RNA is produced by transcription of an expression construct.

35. The method of claim 35 in which the organism is a nematode and the expression construct is contained in a plant, and disease associated with nematode infection of the plant is thereby reduced.

36. A cell containing an expression construct,

wherein the expression construct transcribes at least one ribonucleic acid (RNA) and the RNA forms a double-stranded structure with duplexed strands of ribonucleic acid,

whereby said cell contains the double-stranded RNA structure and is able to inhibit expression of a target gene when the RNA is contacted with an organism containing the target gene.

37. A transgenic animal containing said cell of claim 36.

38. A transgenic plant containing said cell of claim 36.

39. A kit comprising reagents for inhibiting expression of a target gene in a cell,

wherein said kit comprises a means for introduction of a ribonucleic acid (RNA) into the cell in an amount sufficient to inhibit expression of the target gene, and

wherein the RNA has a double-stranded structure with an identical nucleotide sequence as compared to a portion of the target gene.

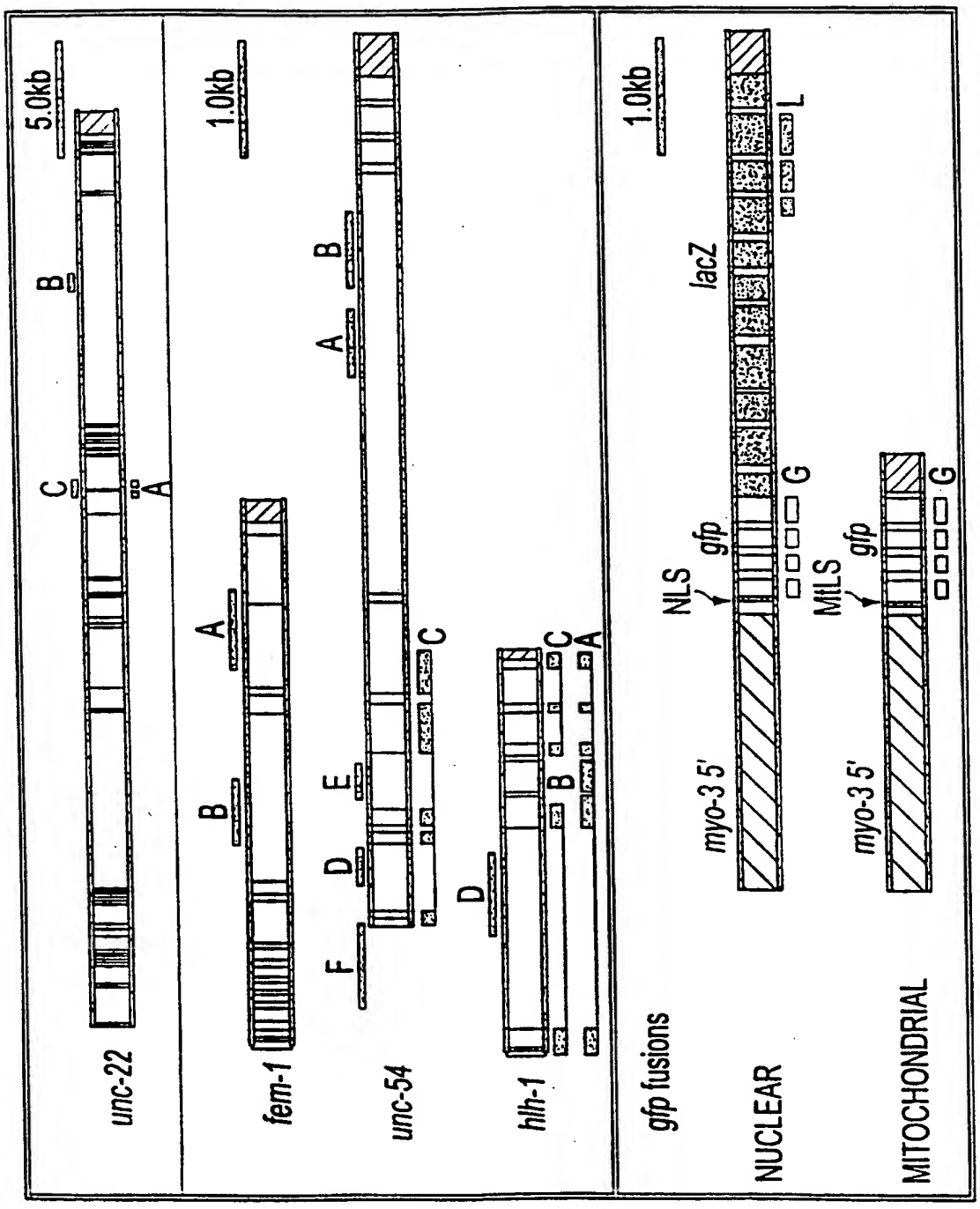

5.0kb
B
C
A
unc-22
1.0kb
A
B
fem-1
A
B
F
D
E
C
unc-54
D
C
A
B
hlh-1
1.0kb
gfp fusions
NLS
gfp
lacZ
L
myo-3 5'
G
NUCLEAR
MLS
gfp
myo-3 5'
G
MITOCHONDRIAL

FIG. 1

FIG. 2G

FIG. 2H

FIG. 2I

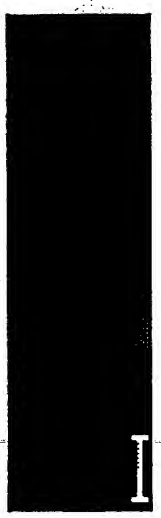


FIG. 2D

FIG. 2E

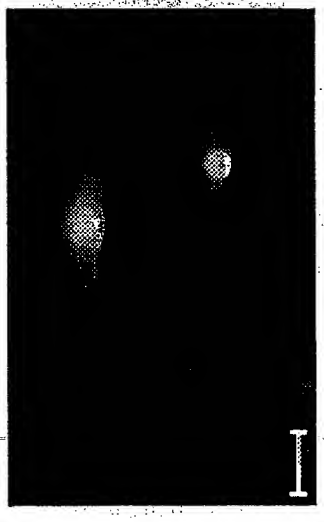


FIG. 2F

FIG. 2A

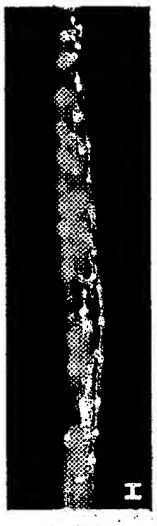


FIG. 2B

FIG. 2C

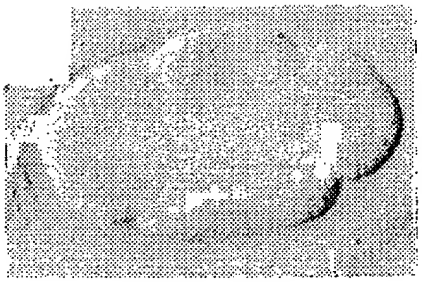


FIG. 3A

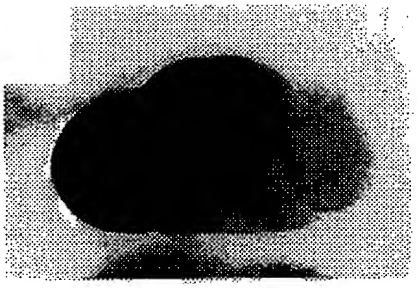


FIG. 3B

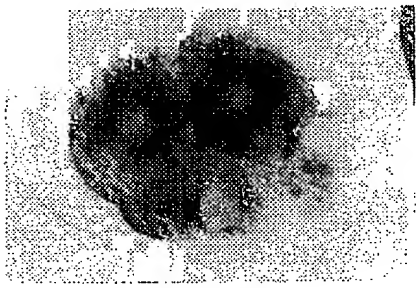


FIG. 3C

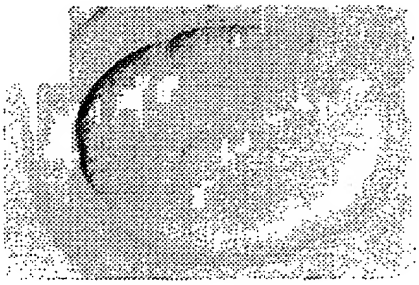


FIG. 3D

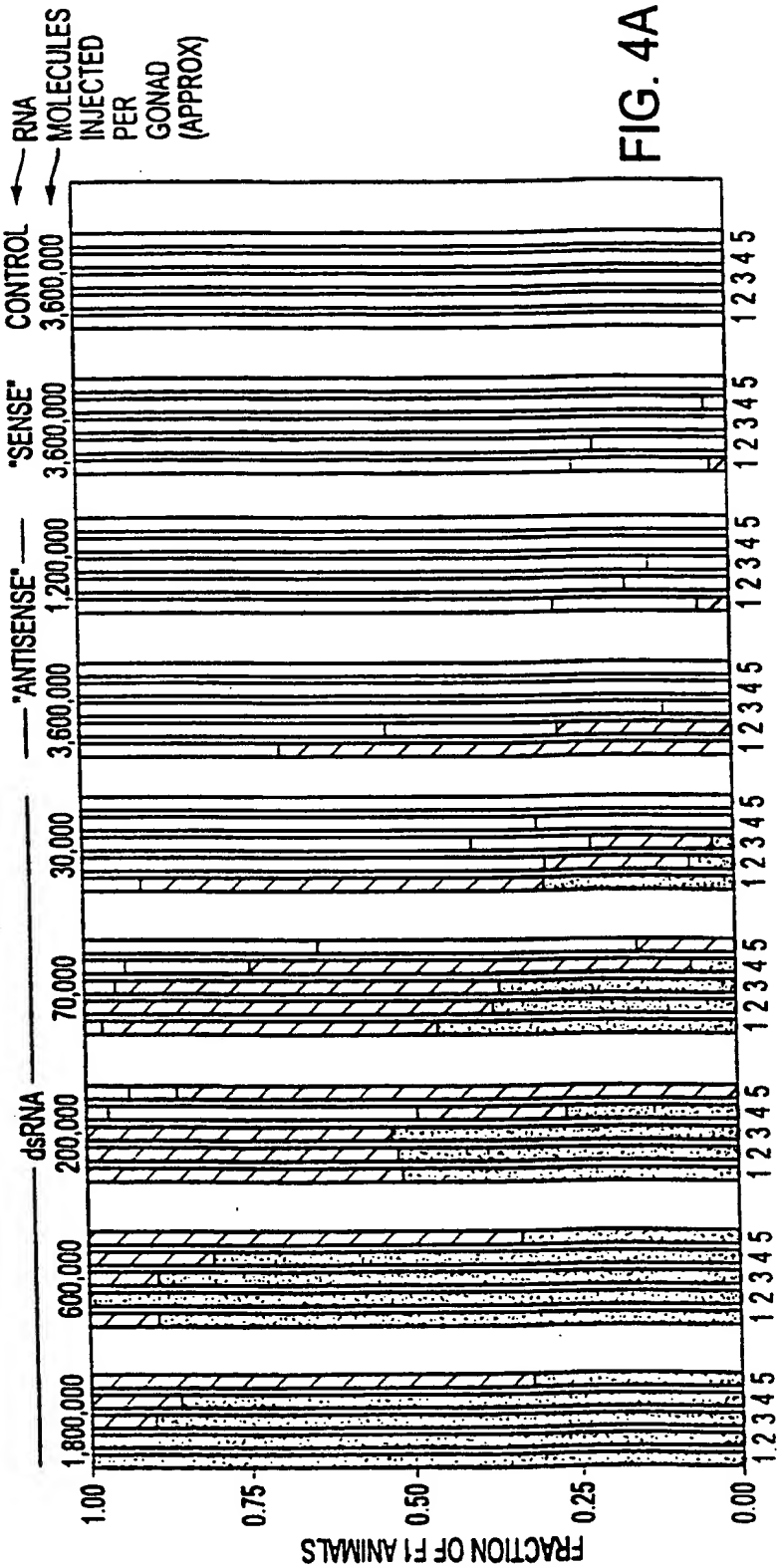

RNA MOLECULES INJECTED PER GONAD (APPROX)
CONTROL
"SENSE"
"ANTISENSE"
dsRNA
3,600,000
3,600,000
1,200,000
3,600,000
30,000
70,000
200,000
600,000
1,800,000
1 2 3 4 5
FRACTION OF F1 ANIMALS
1.00
0.75
0.50
0.25
0.00

FIG. 4A

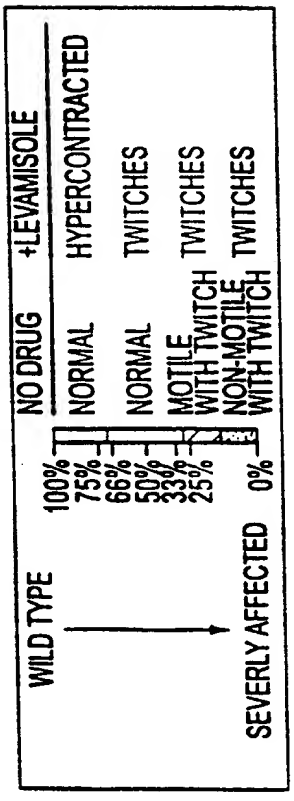

NO DRUG
+LEVAMISOLE
WILD TYPE
100%
75%
66%
50%
33%
25%
0%
SEVERLY AFFECTED
NORMAL
HYPERCONTRACTED
NORMAL
TWITCHES
MOTILE WITH TWITCH
TWITCHES
NON-MOTILE WITH TWITCH
TWITCHES

FIG. 4B

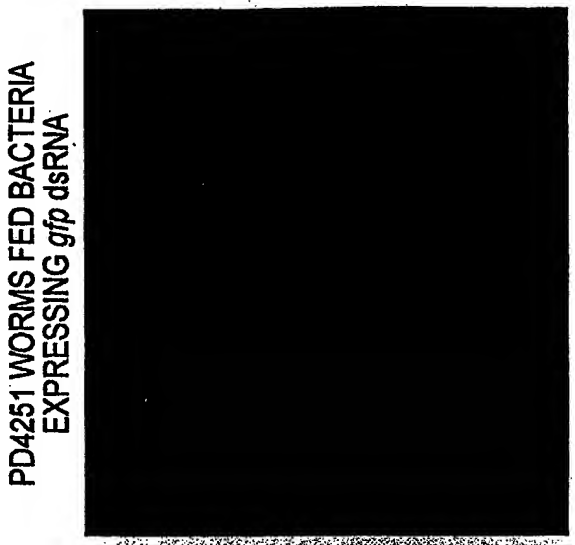


FIG. 5C

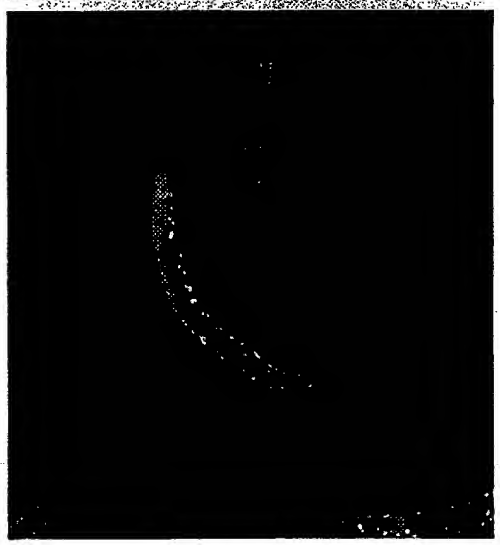


FIG. 5B

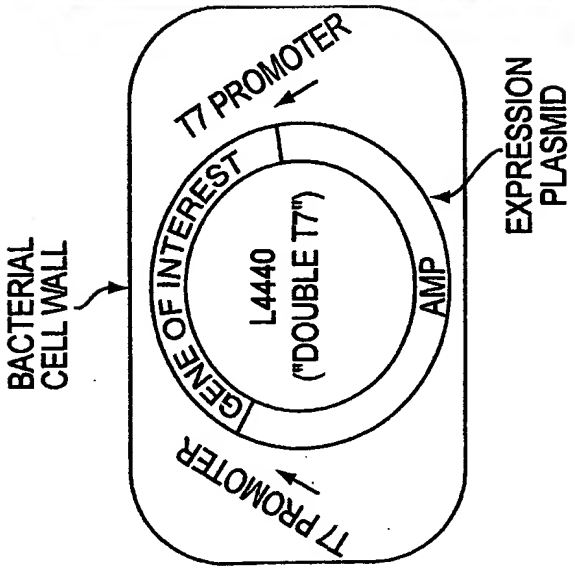


FIG. 5A

Express Mail Label No. EV185285405US
Date Mailed: December 15, 2003

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Atty. Dkt. No. 035718/263003

Kindly acknowledge receipt of the accompanying PROVISIONAL PATENT APPLICATION with Application Transmittal Cover Sheet for: Inventor(s): Gruis et al.

Title of Invention: METHODS AND COMPOSITIONS FOR ALTERING THE FUNCTIONAL PROPERTIES OF SEED STORAGE PROTEINS IN SOYBEAN

Pages of Spec. (including claims and abstract) 66; No. of Claims 1

No. of Drawing Sheets 2 ; Declaration Enclosed NO

Other Documents: Sequence Listing--Paper (21 pages)

Check(s) Enclosed $160

Please place office stamp hereon and Return Postcard to : Nora C. Martinez/KLC

ALSTON & BIRD LLP

17548 U.S. PTO
60/529666

121503

docketing OK 1/8 NM

Attorney's Docket No. 035718/263003 PATENT
Express Mail Label No. EV185285405US

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

Mail Stop Provisional Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 C.F.R. 1.53(c).

INVENTOR(s)

Name: Darren B. Gruis
Residence: DesMoines, IA

Name: Rudolf Jung
Residence: DesMoines, IA

TITLE OF THE INVENTION *(500 characters maximum)*

METHODS AND COMPOSITIONS FOR ALTERING THE FUNCTIONAL PROPERTIES OF SEED STORAGE PROTEINS IN SOYBEAN

CORRESPONDENCE ADDRESS

Kathryn L. Coulter
Registration No. 45,889
Customer Number 29122

Tel Raleigh Office (919) 862-2200
Fax Raleigh Office (919) 862-2260

ENCLOSED APPLICATION PARTS *(check all that apply)*

- ☒ Specification (Number of Pages 66)
- ☒ Drawing(s) (Number of Sheets 2)
- ☐ Application Data Sheet. See 37 CFR 1.76
- ☐ CD(s), Number
- ☒ Other (specify)--Sequence Listing--Paper (21 pages)

METHOD OF PAYMENT OF FILING FEES

- ☐ Applicant claims small entity status
- ☒ Check or money order is enclosed to cover the filing fee.
- ☐ The Commissioner is hereby authorized to charge filing fees to Deposit Account No. 16-0605.
- ☒ Please charge Deposit Account No. 16-0605 for any fee deficiency.

PROVISIONAL FILING FEE AMOUNT(s)
Large Entity $160.00; Small Entity $ 80.00 Filing Fee Amount: $160.00

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

- ☒ No.
- ☐ Yes, the name of the U.S. Government agency and the Government contract number are:

Respectfully submitted,

Kathryn L. Coulter

Kathryn L. Coulter
Registration No. 45,889

Customer No. 00826
ALSTON & BIRD LLP
Bank of America Plaza
101 South Tryon Street, Suite 4000
Charlotte, NC 28280-4000
Tel Raleigh Office (919) 862-2200
Fax Raleigh Office (919) 862-2260

CERTIFICATE OF EXPRESS MAILING

"Express Mail" mailing label number EV185285405US
Date of Deposit: December 15, 2003

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to: Mail Stop Provisional Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

Nora C. Martinez

Nora C. Martinez

EV185285405US

SEQUENCE LISTING

<110> Gruis, Darren
Jung, Rodolf

<120> METHODS AND COMPOSITIONS FOR ALTERING THE FUNCTIONAL PROPERTIES OF SEED STORAGE PROTEINS IN SOYBEAN

<130> 035718/263003

<160> 15

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1769
<212> DNA
<213> Glycine max

<400> 1
gcacgagccc tgttcctgtg tgtgtgagtg accgagtgag tttgtttttc tcagctgata 60
tatatggcgc ttgatcgctc cattataagc aaaacgacgt ggtacagcgt cgtattatgg 120
atgatggtgg tgctggtgag agtgcacggt gcagccgcga ggccgaaccg gaaggagtgg 180
gactcagtca taaagttacc gactgaaccg gtggatgctg actcggatga agtgggaaca 240
cgatgggcgg ttctcgtggc tggttcaaac ggctacggaa actacaggca tcaagcagat 300
gtgtgccatg cgtaccagtt gctgataaaa ggtggactaa aagaagagaa catagtggtg 360
tttatgtacg atgacatagc taccaacgag ttgaatccta gacatggagt catcatcaac 420
caccctgagg gagaagatct gtatgctggt gttcctaagg attacaccgg tgataatgtg 480
acgacggaga acctctttgc tgttattctt ggagacaaga gtaaattgaa gggaggaagt 540
ggcaaagtga tcaacagcaa acccgaggac agaatattta tatactactc tgatcatgga 600
ggtcctggaa tacttgggat gccaaacatg ccataccttt atgccatgga ttttattgat 660
gtcttgaaga agaaacatgc atctggaagt tacaaggaga tggttatata cgtggaagct 720
tgtgaaagtg ggagcgtgtt tgagggtata atgcctaagg atctgaatat ttatgtcaca 780
actgcatcaa atgcacaaga gaatagttgg ggaacttatt gtcctggaat ggatccttct 840
ccacctccag agtacatcac ttgcctaggg gatttgtaca gcgttgcttg gatggaagat 900
agtgaggctc acaatctaaa aagggaatcc gtgaaacaac aatacaaatc ggtaaagcaa 960
cggacttcaa atttcaacaa ctatgcgatg ggttctcatg tgatgcaata tggtgatacc 1020
aacatcacag ctgaaaagct ttatttatac caaggttttg atcctgccac tgtgaacttc 1080
cctccacaaa acggcaggct agaaactaaa atggaagttg ttaaccaaag agatgcagaa 1140
cttttgttca tgtggcaaat gtatcagaga tcaaaccatc agtcagaaaa taagacagac 1200
atcctcaaac aaattgcgga gacagtgaag cataggaaac acatagatgg tagcgtggaa 1260
ttgattggag ttttactgta tggaccagga aaaggttctt ctgttctaca atccgtgagg 1320
gctcctggtt cgtcccttgt tgatgactgg acatgcctaa aatcaatggt tcgggtgttt 1380
gaaactcact gtgggacact gactcagtat ggcatgaaac acatgcgagc attcgccaac 1440
atttgcaaca gtggcgtttc tgaggcctcc atggaagagg cttgtttggc agcctgtgaa 1500
ggctacaatg ctgggctatt gcatccatca aacagaggct acagtgcttg attttgggtt 1560
ttgtacacaa aagctttaaa gcccggttga tgatgtaata tttctctatt gcattctgcc 1620
tactggtttc tgctgcttgt gtcaaatttt ctctaaacta gagtagccca atagcatacg 1680
tgttatgtgc atgtgtcatg tatacaagtg taatactaaa accttctaca taatataaga 1740
ttagttagtt taaaaaaaaa aaaaaaaaa 1769

<210> 2
<211> 495
<212> PRT
<213> Glycine max

<400> 2
Met Ala Leu Asp Arg Ser Ile Ile Ser Lys Thr Thr Trp Tyr Ser Val
1 5 10 15
Val Leu Trp Met Met Val Val Leu Val Arg Val His Gly Ala Ala Ala
20 25 30
Arg Pro Asn Arg Lys Glu Trp Asp Ser Val Ile Lys Leu Pro Thr Glu
35 40 45
Pro Val Asp Ala Asp Ser Asp Glu Val Gly Thr Arg Trp Ala Val Leu
50 55 60
Val Ala Gly Ser Asn Gly Tyr Gly Asn Tyr Arg His Gln Ala Asp Val
65 70 75 80
Cys His Ala Tyr Gln Leu Leu Ile Lys Gly Gly Leu Lys Glu Glu Asn
85 90 95
Ile Val Val Phe Met Tyr Asp Asp Ile Ala Thr Asn Glu Leu Asn Pro
100 105 110
Arg His Gly Val Ile Ile Asn His Pro Glu Gly Glu Asp Leu Tyr Ala
115 120 125
Gly Val Pro Lys Asp Tyr Thr Gly Asp Asn Val Thr Thr Glu Asn Leu
130 135 140
Phe Ala Val Ile Leu Gly Asp Lys Ser Lys Leu Lys Gly Gly Ser Gly
145 150 155 160
Lys Val Ile Asn Ser Lys Pro Glu Asp Arg Ile Phe Ile Tyr Tyr Ser
165 170 175
Asp His Gly Gly Pro Gly Ile Leu Gly Met Pro Asn Met Pro Tyr Leu
180 185 190
Tyr Ala Met Asp Phe Ile Asp Val Leu Lys Lys Lys His Ala Ser Gly
195 200 205
Ser Tyr Lys Glu Met Val Ile Tyr Val Glu Ala Cys Glu Ser Gly Ser
210 215 220
Val Phe Glu Gly Ile Met Pro Lys Asp Leu Asn Ile Tyr Val Thr Thr
225 230 235 240
Ala Ser Asn Ala Gln Glu Asn Ser Trp Gly Thr Tyr Cys Pro Gly Met
245 250 255
Asp Pro Ser Pro Pro Pro Glu Tyr Ile Thr Cys Leu Gly Asp Leu Tyr
260 265 270
Ser Val Ala Trp Met Glu Asp Ser Glu Ala His Asn Leu Lys Arg Glu
275 280 285
Ser Val Lys Gln Gln Tyr Lys Ser Val Lys Gln Arg Thr Ser Asn Phe
290 295 300
Asn Asn Tyr Ala Met Gly Ser His Val Met Gln Tyr Gly Asp Thr Asn
305 310 315 320
Ile Thr Ala Glu Lys Leu Tyr Leu Tyr Gln Gly Phe Asp Pro Ala Thr
325 330 335
Val Asn Phe Pro Pro Gln Asn Gly Arg Leu Glu Thr Lys Met Glu Val
340 345 350
Val Asn Gln Arg Asp Ala Glu Leu Leu Phe Met Trp Gln Met Tyr Gln
355 360 365
Arg Ser Asn His Gln Ser Glu Asn Lys Thr Asp Ile Leu Lys Gln Ile
370 375 380
Ala Glu Thr Val Lys His Arg Lys His Ile Asp Gly Ser Val Glu Leu
385 390 395 400
Ile Gly Val Leu Leu Tyr Gly Pro Gly Lys Gly Ser Ser Val Leu Gln
405 410 415
Ser Val Arg Ala Pro Gly Ser Ser Leu Val Asp Asp Trp Thr Cys Leu
420 425 430
Lys Ser Met Val Arg Val Phe Glu Thr His Cys Gly Thr Leu Thr Gln

```
        435                 440                 445
Tyr Gly Met Lys His Met Arg Ala Phe Ala Asn Ile Cys Asn Ser Gly
    450                 455                 460
Val Ser Glu Ala Ser Met Glu Glu Ala Cys Leu Ala Ala Cys Glu Gly
465                 470                 475                 480
Tyr Asn Ala Gly Leu Leu His Pro Ser Asn Arg Gly Tyr Ser Ala
                485                 490                 495
```

<210> 3
<211> 1806
<212> DNA
<213> Glycine max

<400> 3

```
gcacgaggtg agtctttctt agctgatatg gcggttgatc gctcccttac gaggtgctgt 60
agcctcgtac tgtggtcgtg gatgttgctg aggatgatga tggcgcaggg tgcagccgcg 120
agggccaacc ggaaggagtg ggactcggtc ataaagttac cggctgaacc ggtcgatgct 180
gactcggatc atgaagtggg aacacgatgg gcggttcttg tggctggttc aaacggctat 240
ggaaactaca ggcatcaagc agatgtgtgc catgcgtacc agttgctgat aaaaggtggg 300
ctaaaagaag agaacatagt ggtgtttatg tacgatgaca tagctacaga cgagttaaat 360
cccagacctg gagtcatcat caaccaccct gagggacaag atgtgtatgc tggtgttcct 420
aaggattaca ccggtgagaa tgtgacggcc cagaacctct ttgccgttat tcttggagac 480
aagaataaag tgaagggagg aagtggcaaa gtgatcaata gcaaacctga ggacagaata 540
tttatatact actctgatca tggaggtccg ggagttcttg ggatgccaaa catgccatac 600
ctttatgcta tggactttat tgaagtcttg aagaagaaac atgcatctgg aggttacaag 660
aagatggtca tatacgtgga agcttgtgaa agtgggagca tgtttgaggg tataatgcct 720
aaggatctgc agatttatgt cacaactgca tccaatgcac aagagaatag ttggggaact 780
tattgtcctg gaatggatcc ttctccacct ccagagtaca tcacttgcct aggggatttg 840
tacagtgttg cttggatgga agatagtgag actcataatc taaaaaggga gtccgtgaaa 900
caacaataca aatcggtaaa gcaacggact tcaaatttca acaactatgc gatgggttct 960
catgtgatgc aatacggtga cacaaacatc acagctgaaa agctttattt ataccaaggt 1020
tttgatcctg ccgctgtgaa cttccctcca cagaacggaa ggctagaaac taaaatggaa 1080
gttgttaacc aaagagatgc agaacttttc ttcatgtggc aaatgtatca gagatcaaac 1140
catcagccag aaaagaagac agacatcctc aaacagatag cggagacagt gaagcatagg 1200
aaacacatag atggtagcgt ggaattgatt ggagttttat tgtatggacc aggaaaaggt 1260
tcttctgttc tacaatccat gagggctcct ggtcttgccc ttgttgatga ctggacatgc 1320
ctaaaatcaa tggttcgggt gtttgagact cactgtggga cactgactca gtatggcatg 1380
aaacacatgc gagcatttgc caacatttgc aacagcggtg tttctgaggc ctccatggaa 1440
gaggtttgtg tggcagcttg tgaaggctac gattctgggc tattacatcc atcaaacaaa 1500
ggctatagtg cttgattttg ggttttgtac acagcttaaa aacccggttg atgatgtaat 1560
acttctctat tgcattctcc ctactggttt ctgctgcatg tgtcaaattt tctctaaact 1620
agagtagccc aatagcatac gtgttatgag cattggtcat gtatataagt gtaatagtaa 1680
tatcttttac atattataag atcagttagt ttggtttact agtgtctgtt tcaagctcta 1740
ttttcttgaa ctcaactcct tctaaatcaa ggagattttt cttaaaaaaa aaaaaaaaaa 1800
aaaaaa                                                          1806
```

<210> 4
<211> 495
<212> PRT
<213> Glycine max

<400> 4

```
Met Ala Val Asp Arg Ser Leu Thr Arg Cys Cys Ser Leu Val Leu Trp
 1               5                  10                  15
Ser Trp Met Leu Leu Arg Met Met Met Ala Gln Gly Ala Ala Ala Arg
            20                  25                  30
```

```
Ala Asn Arg Lys Glu Trp Asp Ser Val Ile Lys Leu Pro Ala Glu Pro
        35                  40                  45
Val Asp Ala Asp Ser Asp His Glu Val Gly Thr Arg Trp Ala Val Leu
    50                  55                  60
Val Ala Gly Ser Asn Gly Tyr Gly Asn Tyr Arg His Gln Ala Asp Val
65                  70                  75                  80
Cys His Ala Tyr Gln Leu Leu Ile Lys Gly Gly Leu Lys Glu Glu Asn
                85                  90                  95
Ile Val Val Phe Met Tyr Asp Asp Ile Ala Thr Asp Glu Leu Asn Pro
            100                 105                 110
Arg Pro Gly Val Ile Ile Asn His Pro Glu Gly Gln Asp Val Tyr Ala
        115                 120                 125
Gly Val Pro Lys Asp Tyr Thr Gly Glu Asn Val Thr Ala Gln Asn Leu
    130                 135                 140
Phe Ala Val Ile Leu Gly Asp Lys Asn Lys Val Lys Gly Gly Ser Gly
145                 150                 155                 160
Lys Val Ile Asn Ser Lys Pro Glu Asp Arg Ile Phe Ile Tyr Tyr Ser
                165                 170                 175
Asp His Gly Gly Pro Gly Val Leu Gly Met Pro Asn Met Pro Tyr Leu
            180                 185                 190
Tyr Ala Met Asp Phe Ile Glu Val Leu Lys Lys Lys His Ala Ser Gly
        195                 200                 205
Gly Tyr Lys Lys Met Val Ile Tyr Val Glu Ala Cys Glu Ser Gly Ser
    210                 215                 220
Met Phe Glu Gly Ile Met Pro Lys Asp Leu Gln Ile Tyr Val Thr Thr
225                 230                 235                 240
Ala Ser Asn Ala Gln Glu Asn Ser Trp Gly Thr Tyr Cys Pro Gly Met
                245                 250                 255
Asp Pro Ser Pro Pro Pro Glu Tyr Ile Thr Cys Leu Gly Asp Leu Tyr
            260                 265                 270
Ser Val Ala Trp Met Glu Asp Ser Glu Thr His Asn Leu Lys Arg Glu
        275                 280                 285
Ser Val Lys Gln Gln Tyr Lys Ser Val Lys Gln Arg Thr Ser Asn Phe
    290                 295                 300
Asn Asn Tyr Ala Met Gly Ser His Val Met Gln Tyr Gly Asp Thr Asn
305                 310                 315                 320
Ile Thr Ala Glu Lys Leu Tyr Leu Tyr Gln Gly Phe Asp Pro Ala Ala
                325                 330                 335
Val Asn Phe Pro Pro Gln Asn Gly Arg Leu Glu Thr Lys Met Glu Val
            340                 345                 350
Val Asn Gln Arg Asp Ala Glu Leu Phe Phe Met Trp Gln Met Tyr Gln
        355                 360                 365
Arg Ser Asn His Gln Pro Glu Lys Lys Thr Asp Ile Leu Lys Gln Ile
    370                 375                 380
Ala Glu Thr Val Lys His Arg Lys His Ile Asp Gly Ser Val Glu Leu
385                 390                 395                 400
Ile Gly Val Leu Leu Tyr Gly Pro Gly Lys Gly Ser Ser Val Leu Gln
                405                 410                 415
Ser Met Arg Ala Pro Gly Leu Ala Leu Val Asp Asp Trp Thr Cys Leu
            420                 425                 430
Lys Ser Met Val Arg Val Phe Glu Thr His Cys Gly Thr Leu Thr Gln
        435                 440                 445
Tyr Gly Met Lys His Met Arg Ala Phe Ala Asn Ile Cys Asn Ser Gly
    450                 455                 460
Val Ser Glu Ala Ser Met Glu Glu Val Cys Val Ala Ala Cys Glu Gly
465                 470                 475                 480
Tyr Asp Ser Gly Leu Leu His Pro Ser Asn Lys Gly Tyr Ser Ala
```

485 490 495

<210> 5
<211> 1936
<212> DNA
<213> Glycine max

<400> 5
gcacgagaat taaattaata gaggatgaaa ttctagttta aggaaggttg gttggttggg 60
tgggggtagg agatactctc attcacctcc catcatcatt ataatcattc attccaacct 120
acccttattc ttcttcttca atttcacacc catcatggac cgttttccga tcctctttct 180
cgtcgccacc ctcatcaccc tcgcctccgg tgcccgccac gatattctcc ggttaccctc 240
cgaagcttcc aggttcttca aagcacctgc taatgccgat caaaacgatg agggcaccag 300
gtgggccgtt ttagttgccg gttccaatgg ctactggaat tacaggcacc agtctgatgt 360
ttgccatgca tatcaactac tgaggaaagg tggtgtgaaa gaggaaaata ttgttgtatt 420
tatgtatgat gacattgctt tcaatgaaga gaacccacgg cctggagtca ttattaacag 480
tccacacgga aatgatgttt acaagggagt tcctaaggat tacgttggtg aagatgttac 540
tgttgacaac ttttttgctg ctatacttgg aaataagtca gctcttactg gtggcagtgg 600
gaaggttgtg gatagtggcc ccaatgatca tatatttata tactactctg atcatggcgg 660
tccgggagtg ctagggatgc ctactaatcc atacatgtat gcatccgatc tgattgaagt 720
cttgaagaag aagcatgctt ctggaactta taaaagccta gtattttatc tagaggcatg 780
tgaatctggg agtatctttg aaggtcttct tccagaaggt ctgaatatct atgcaacaac 840
agcttcaaat gctgaagaaa gcagttgggg aacatattgt cctggggagt atcctagtcc 900
tccccctgaa tatgaaacct gcctgggtga cctgtacagt gttgcttgga tggaagatag 960
tgacatacac aatttgcgaa cagaaacttt acatcaacaa tacgacttgg tcaaagaaag 1020
gactatgaat ggaaattcaa tctatggttc ccacgtgatg cagtatggtg acatagggct 1080
tagcaagaac aatcttgtct tatatttggg tacaaatcct gctaatgata attttacttt 1140
tgtgcataaa aactcattgg tgccaccttc aaaagcagtc aaccaacgtg atgcagatct 1200
catccatttc tgggataagt tccgcaaagc tcctgtgggt tcttctagga aagctgcagc 1260
tgagaaagaa attctggaag caatgtctca cagaatgcat atagatgaca acatgaaact 1320
tattggaaag ctcttatttg gcattgaaaa gggtccagaa ctgcttagca gtgttagacc 1380
tgctgggcaa ccacttgttg atgactggga ctgccttaaa acactggtta ggacttttga 1440
gacacattgt ggatctctgt ctcagtatgg gatgaaacat atgaggtcct ttgcaaactt 1500
ctgcaacgct ggaatacgga aagagcaaat ggctgaggcc tcggcacaag catgtgtcag 1560
tatccctgca agttcctgga gttctctgca caggggtttc agtgcataat tcctagaatc 1620
cgctccattg aagacagagt atagtcgttg taacattatt ctttacgagc gttatgtact 1680
gtacctggac atgatttctt ataccaaccc tgttaataag catgggacgc tggggaaacc 1740
tatttacatt gtaatttcgt gcaaaataga tgctgtaaca aaggcatttt acttttactt 1800
ggggagaggc agtggaacca taaggacctt ggaaattctg attaatatga cagggcacaa 1860
tatcgtgttt gtaagccaac gctttatttt tattttatgg taaccccttt ctgtggataa 1920
aaaaaaaaaa aaaaaa 1936

<210> 6
<211> 484
<212> PRT
<213> Glycine max

<400> 6
Met Asp Arg Phe Pro Ile Leu Phe Leu Val Ala Thr Leu Ile Thr Leu
1 5 10 15
Ala Ser Gly Ala Arg His Asp Ile Leu Arg Leu Pro Ser Glu Ala Ser
20 25 30
Arg Phe Phe Lys Ala Pro Ala Asn Ala Asp Gln Asn Asp Glu Gly Thr
35 40 45
Arg Trp Ala Val Leu Val Ala Gly Ser Asn Gly Tyr Trp Asn Tyr Arg
50 55 60

```
His Gln Ser Asp Val Cys His Ala Tyr Gln Leu Leu Arg Lys Gly Gly
65                  70                  75                  80
Val Lys Glu Glu Asn Ile Val Val Phe Met Tyr Asp Asp Ile Ala Phe
                85                  90                  95
Asn Glu Glu Asn Pro Arg Pro Gly Val Ile Ile Asn Ser Pro His Gly
            100                 105                 110
Asn Asp Val Tyr Lys Gly Val Pro Lys Asp Tyr Val Gly Glu Asp Val
        115                 120                 125
Thr Val Asp Asn Phe Phe Ala Ala Ile Leu Gly Asn Lys Ser Ala Leu
    130                 135                 140
Thr Gly Gly Ser Gly Lys Val Val Asp Ser Gly Pro Asn Asp His Ile
145                 150                 155                 160
Phe Ile Tyr Tyr Ser Asp His Gly Gly Pro Gly Val Leu Gly Met Pro
                165                 170                 175
Thr Asn Pro Tyr Met Tyr Ala Ser Asp Leu Ile Glu Val Leu Lys Lys
            180                 185                 190
Lys His Ala Ser Gly Thr Tyr Lys Ser Leu Val Phe Tyr Leu Glu Ala
        195                 200                 205
Cys Glu Ser Gly Ser Ile Phe Glu Gly Leu Leu Pro Glu Gly Leu Asn
    210                 215                 220
Ile Tyr Ala Thr Thr Ala Ser Asn Ala Glu Glu Ser Ser Trp Gly Thr
225                 230                 235                 240
Tyr Cys Pro Gly Glu Tyr Pro Ser Pro Pro Pro Glu Tyr Glu Thr Cys
                245                 250                 255
Leu Gly Asp Leu Tyr Ser Val Ala Trp Met Glu Asp Ser Asp Ile His
            260                 265                 270
Asn Leu Arg Thr Glu Thr Leu His Gln Gln Tyr Asp Leu Val Lys Glu
        275                 280                 285
Arg Thr Met Asn Gly Asn Ser Ile Tyr Gly Ser His Val Met Gln Tyr
    290                 295                 300
Gly Asp Ile Gly Leu Ser Lys Asn Asn Leu Val Leu Tyr Leu Gly Thr
305                 310                 315                 320
Asn Pro Ala Asn Asp Asn Phe Thr Phe Val His Lys Asn Ser Leu Val
                325                 330                 335
Pro Pro Ser Lys Ala Val Asn Gln Arg Asp Ala Asp Leu Ile His Phe
            340                 345                 350
Trp Asp Lys Phe Arg Lys Ala Pro Val Gly Ser Ser Arg Lys Ala Ala
        355                 360                 365
Ala Glu Lys Glu Ile Leu Glu Ala Met Ser His Arg Met His Ile Asp
    370                 375                 380
Asp Asn Met Lys Leu Ile Gly Lys Leu Leu Phe Gly Ile Glu Lys Gly
385                 390                 395                 400
Pro Glu Leu Leu Ser Ser Val Arg Pro Ala Gly Gln Pro Leu Val Asp
                405                 410                 415
Asp Trp Asp Cys Leu Lys Thr Leu Val Arg Thr Phe Glu Thr His Cys
            420                 425                 430
Gly Ser Leu Ser Gln Tyr Gly Met Lys His Met Arg Ser Phe Ala Asn
        435                 440                 445
Phe Cys Asn Ala Gly Ile Arg Lys Glu Gln Met Ala Glu Ala Ser Ala
    450                 455                 460
Gln Ala Cys Val Ser Ile Pro Ala Ser Ser Trp Ser Ser Leu His Arg
465                 470                 475                 480
Gly Phe Ser Ala
```

<210> 7

<211> 1942
<212> DNA
<213> Glycine max

<400> 7
gcacgagctc tctctctctc tctctctctc tctctctctc tctctctctc tctctctctc 60
tctctctctc tctctctctc tctctctctc tctctctctc tctcctcact cgttcattcc 120
aacctaccct tattcttctt cttcaattcc acacccatca tggaccgttt tccgatcctc 180
tttctcctcg ccaccctcat caccctcgcc tccggtgccc gccacgatat tctccggtta 240
ccctccgaag catccacttt tttcaaagca cccggtggcg atcaaaacga tgagggcacg 300
aggtgggccg ttttaattgc cggttccaat ggctactgga attacaggca ccagtctgat 360
gtttgccatg cgtatcaact actgaggaaa ggtggtctca aagaagaaaa tattgttgta 420
tttatgtatg atgacattgc tttcaacgaa gagaacccgc gacctggagt cattattaac 480
agtccacatg gaaatgatgt ttacaaggga gtccctaagg attacattgg tgaagatgta 540
actgttggca actttttttgc tgctatactt ggaaataagt cagctcttac tggtggcagt 600
gggaaggttg tggatagtgg tcccaatgat catatattta tatattactc tgatcatggc 660
ggtcctggag tgctagggat gcctactaat ccatacatgt atgcatctga tctgattgaa 720
gtcttgaaga agaagcatgc ttctggaagt tataaaagcc tagtatttta tctagaggca 780
tgtgaatctg ggagtatctt tgaaggtctt cttcctgaag gtctgaatat ctatgcaaca 840
acagcttcaa atgcagaaga aagcagttgg ggaacatatt gtcctgggga gtatcctagt 900
cctccctctg aatatgaaac ctgcctgggt gacctgtaca gtgttgcttg gatggaagac 960
agtgacatac acaatttgca aacagaaact ttacatcaac aatacgaatt ggtcaaacaa 1020
aggactatga atggaaattc aatttatggt tcccacgtga tgcagtatgg tgacataggg 1080
cttagcgaga acaatctcgt cttatatttg ggtacaaatc ctgctaatga taattttact 1140
tttgtgctta aaaactcatt ggtgccacct tcaaaagcag tcaaccaacg tgatgcagat 1200
ctcatccatt tttgggataa gttccgcaaa gctcctgtgg gttcttctag gaaagctgca 1260
gctgagaaac aaattcttga agcaatgtct cacagaatgc atatagatga cagcatgaaa 1320
cgtattggaa agctcttctt tggcattgaa aagggtccag aactgcttag cagtgttaga 1380
cctgctgggc aaccacttgt tgatgactgg gactgcctta aaacattggt taggactttt 1440
gagacacatt gtggatccct gtctcagtat gggatgaaac atatgaggtc ctttgcaaac 1500
ttctgcaacg ctggaatacg aaaagagcaa atggctgagg cctcagcaca agcatgtgtc 1560
aatatccctg ctagttcctg gagttctatg cacaggggtt tcagtgcata attcctagaa 1620
tgcgctccat tgaagaccga gtatagtcgt tgtaacatta ttctttacga gtgttatgga 1680
ctgtactctc tgctcatgat ttcttatacc aaccctgtaa atacaaatgg gacgctgggg 1740
aaacctcttt acattatagt ttcctgcaaa atagatgctg taacaaagac attttacttt 1800
tacttgggga gaggcagtgg aaccataagg acccttggaa cttctaatta atacgacagg 1860
gcacaatacc gtgtttgtaa gccaacgctt tgtttcaatt taatggtaac cccgttgtgt 1920
agaaaaaaaa aaaaaaaaaa aa 1942

<210> 8
<211> 483
<212> PRT
<213> Glycine max

<400> 8
Met Asp Arg Phe Pro Ile Leu Phe Leu Leu Ala Thr Leu Ile Thr Leu
 1   5   10   15
Ala Ser Gly Ala Arg His Asp Ile Leu Arg Leu Pro Ser Glu Ala Ser
            20   25   30
Thr Phe Phe Lys Ala Pro Gly Gly Asp Gln Asn Asp Glu Gly Thr Arg
        35   40   45
Trp Ala Val Leu Ile Ala Gly Ser Asn Gly Tyr Trp Asn Tyr Arg His
    50   55   60
Gln Ser Asp Val Cys His Ala Tyr Gln Leu Leu Arg Lys Gly Gly Leu
65   70   75   80
Lys Glu Glu Asn Ile Val Val Phe Met Tyr Asp Asp Ile Ala Phe Asn
                85   90   95

```
Glu Glu Asn Pro Arg Pro Gly Val Ile Ile Asn Ser Pro His Gly Asn
            100                 105                 110
Asp Val Tyr Lys Gly Val Pro Lys Asp Tyr Ile Gly Glu Asp Val Thr
        115                 120                 125
Val Gly Asn Phe Phe Ala Ala Ile Leu Gly Asn Lys Ser Ala Leu Thr
    130                 135                 140
Gly Gly Ser Gly Lys Val Val Asp Ser Gly Pro Asn Asp His Ile Phe
145                 150                 155                 160
Ile Tyr Tyr Ser Asp His Gly Gly Pro Gly Val Leu Gly Met Pro Thr
                165                 170                 175
Asn Pro Tyr Met Tyr Ala Ser Asp Leu Ile Glu Val Leu Lys Lys Lys
            180                 185                 190
His Ala Ser Gly Ser Tyr Lys Ser Leu Val Phe Tyr Leu Glu Ala Cys
        195                 200                 205
Glu Ser Gly Ser Ile Phe Glu Gly Leu Leu Pro Glu Gly Leu Asn Ile
    210                 215                 220
Tyr Ala Thr Thr Ala Ser Asn Ala Glu Glu Ser Ser Trp Gly Thr Tyr
225                 230                 235                 240
Cys Pro Gly Glu Tyr Pro Ser Pro Pro Ser Glu Tyr Glu Thr Cys Leu
                245                 250                 255
Gly Asp Leu Tyr Ser Val Ala Trp Met Glu Asp Ser Asp Ile His Asn
            260                 265                 270
Leu Gln Thr Glu Thr Leu His Gln Gln Tyr Glu Leu Val Lys Gln Arg
        275                 280                 285
Thr Met Asn Gly Asn Ser Ile Tyr Gly Ser His Val Met Gln Tyr Gly
    290                 295                 300
Asp Ile Gly Leu Ser Glu Asn Asn Leu Val Leu Tyr Leu Gly Thr Asn
305                 310                 315                 320
Pro Ala Asn Asp Asn Phe Thr Phe Val Leu Lys Asn Ser Leu Val Pro
                325                 330                 335
Pro Ser Lys Ala Val Asn Gln Arg Asp Ala Asp Leu Ile His Phe Trp
            340                 345                 350
Asp Lys Phe Arg Lys Ala Pro Val Gly Ser Ser Arg Lys Ala Ala Ala
        355                 360                 365
Glu Lys Gln Ile Leu Glu Ala Met Ser His Arg Met His Ile Asp Asp
    370                 375                 380
Ser Met Lys Arg Ile Gly Lys Leu Phe Phe Gly Ile Glu Lys Gly Pro
385                 390                 395                 400
Glu Leu Leu Ser Ser Val Arg Pro Ala Gly Gln Pro Leu Val Asp Asp
                405                 410                 415
Trp Asp Cys Leu Lys Thr Leu Val Arg Thr Phe Glu Thr His Cys Gly
            420                 425                 430
Ser Leu Ser Gln Tyr Gly Met Lys His Met Arg Ser Phe Ala Asn Phe
        435                 440                 445
Cys Asn Ala Gly Ile Arg Lys Glu Gln Met Ala Glu Ala Ser Ala Gln
    450                 455                 460
Ala Cys Val Asn Ile Pro Ala Ser Ser Trp Ser Ser Met His Arg Gly
465                 470                 475                 480
Phe Ser Ala
```

<210> 9
<211> 1948
<212> DNA
<213> Glycine max

<400> 9
gcaccagaaa atgcccactt tttttcttcc aacgctcctc ctccttctca tagccttcgc 60
cacctctgtc tccggccgcc gtgacctcgt cggagacttt ctccggctgc cctccgaaac 120
tgataacgac gacaacttca agggcacccg gtgggccgtc ctcctcgccg gttccaatgg 180
ttactggaat tacagacatc aggctgatgt ttgtcacgcc tatcaaatat tgaggaaagg 240
tggtctgaaa gaagaaaata ttattgtttt tatgtatgat gacattgcat tcaatgggga 300
aaacccaagg cctggagtca tcattaacaa accagatgga ggtgatgttt ataaaggagt 360
tccaaaggat tacaccggcg aagatgttac tgttgataac ttttttgctg ctttacttgg 420
aaataagtca gcactgactg gtggcagtgg gaaggttgtg gacagtggtc ctgatgatca 480
tatatttgta tactatactg accatggagg tcctggggtg ctcgggatgc ctgctggtcc 540
ttacttatac gcggatgatc tgattgaagt cttgaagaaa aagcatgctt ctggaacata 600
taaaaaccta gtattttatc tggaggcatg tgaatctggg agtatctttg aaggtcttct 660
tcctgaagat atcaatattt atgcaaccac tgcttccaat gcagaagaaa gtagttgggg 720
aacatattgc cccggggagt atcctagtcc tcccccagaa tatacaacct gtttgggtga 780
cttgtacagt gttgcttgga tggaagacag tgacagacac aatttgcgaa cagaaactct 840
gcaccaacaa tataaattgg ttaaagagag gactatatct ggagattcat actatggctc 900
tcacgtgatg cagtatggtg atgtagggct tagcagagat gttctcttcc attatttggg 960
tacagatcct gctaatgata atttcacttt tgtggatgaa aactccttat ggtcaccttc 1020
aaaaccagtc aaccaacgtg atgctgatct catccatttt tgggataagt tccgcaaagc 1080
tcctgagggt tctctcagga aaaatacagc tcagaaacaa gttttggaag caatgtctca 1140
cagaatgcat gtagacaaca gtgtaaaact gattgggaag cttttatttg gcattgaaaa 1200
gggtccagaa gtactcaacg ctgttagacc ggctggatcg gcacttgttg atgactggca 1260
ctgcctgaaa accatggtga ggacttttga gacacattgt ggatccttgt ctcaatacgg 1320
gatgaaacac atgaggtcct ttgcaaacat ctgcaatgta gggataaaga atgaacaaat 1380
ggctgaggct tcagcacaag cttgtgtcag tattccttcc aatccctgga gttctctgca 1440
aaggggtttc agtgcataat aactccctgt aatgtgcact agtaaagacc aaagtatgat 1500
tattgttaca ttatgttaca tggttgtact tgtatataca tatcttgtcc cacctttgta 1560
aatacaattg ggacactact aggattggga agaagggtct ttacatttat agtttggcaa 1620
atagatattg caactacctt tgtataattc tatttctgaa gaagcaatta caatttacaa 1680
gggatggtgc catttacggc ataaggatta aggagggata aagggaccaa ttgctttgga 1740
atatccactc attacaatgc atgtatgaca acacatagta atatgatgtg tgtttttatt 1800
cagtgggcaa ctggcagatc gggttttccc tggtcacttt tgtataatta ttccggaaga 1860
atttatgatg ccaaaattat tgtttaatat taatgacaac ttgtatttat ttttgtaaaa 1920
aaaaaaaaaa aaaaaaaaaa aaaaaaaa 1948

<210> 10
<211> 482
<212> PRT
<213> Glycine max

<400> 10
Met Pro Thr Phe Phe Leu Pro Thr Leu Leu Leu Leu Leu Ile Ala Phe
 1   5   10   15
Ala Thr Ser Val Ser Gly Arg Arg Asp Leu Val Gly Asp Phe Leu Arg
            20   25   30
Leu Pro Ser Glu Thr Asp Asn Asp Asp Asn Phe Lys Gly Thr Arg Trp
        35   40   45
Ala Val Leu Leu Ala Gly Ser Asn Gly Tyr Trp Asn Tyr Arg His Gln
    50   55   60
Ala Asp Val Cys His Ala Tyr Gln Ile Leu Arg Lys Gly Gly Leu Lys
65   70   75   80
Glu Glu Asn Ile Ile Val Phe Met Tyr Asp Asp Ile Ala Phe Asn Gly
                85   90   95
Glu Asn Pro Arg Pro Gly Val Ile Ile Asn Lys Pro Asp Gly Gly Asp
            100   105   110
Val Tyr Lys Gly Val Pro Lys Asp Tyr Thr Gly Glu Asp Val Thr Val
        115   120   125

```
Asp Asn Phe Phe Ala Ala Leu Leu Gly Asn Lys Ser Ala Leu Thr Gly
    130                 135                 140
Gly Ser Gly Lys Val Val Asp Ser Gly Pro Asp Asp His Ile Phe Val
145                 150                 155                 160
Tyr Tyr Thr Asp His Gly Gly Pro Gly Val Leu Gly Met Pro Ala Gly
                165                 170                 175
Pro Tyr Leu Tyr Ala Asp Asp Leu Ile Glu Val Leu Lys Lys Lys His
            180                 185                 190
Ala Ser Gly Thr Tyr Lys Asn Leu Val Phe Tyr Leu Glu Ala Cys Glu
        195                 200                 205
Ser Gly Ser Ile Phe Glu Gly Leu Leu Pro Glu Asp Ile Asn Ile Tyr
    210                 215                 220
Ala Thr Thr Ala Ser Asn Ala Glu Glu Ser Ser Trp Gly Thr Tyr Cys
225                 230                 235                 240
Pro Gly Glu Tyr Pro Ser Pro Pro Pro Glu Tyr Thr Thr Cys Leu Gly
                245                 250                 255
Asp Leu Tyr Ser Val Ala Trp Met Glu Asp Ser Asp Arg His Asn Leu
            260                 265                 270
Arg Thr Glu Thr Leu His Gln Gln Tyr Lys Leu Val Lys Glu Arg Thr
        275                 280                 285
Ile Ser Gly Asp Ser Tyr Tyr Gly Ser His Val Met Gln Tyr Gly Asp
    290                 295                 300
Val Gly Leu Ser Arg Asp Val Leu Phe His Tyr Leu Gly Thr Asp Pro
305                 310                 315                 320
Ala Asn Asp Asn Phe Thr Phe Val Asp Glu Asn Ser Leu Trp Ser Pro
                325                 330                 335
Ser Lys Pro Val Asn Gln Arg Asp Ala Asp Leu Ile His Phe Trp Asp
            340                 345                 350
Lys Phe Arg Lys Ala Pro Glu Gly Ser Leu Arg Lys Asn Thr Ala Gln
        355                 360                 365
Lys Gln Val Leu Glu Ala Met Ser His Arg Met His Val Asp Asn Ser
    370                 375                 380
Val Lys Leu Ile Gly Lys Leu Leu Phe Gly Ile Glu Lys Gly Pro Glu
385                 390                 395                 400
Val Leu Asn Ala Val Arg Pro Ala Gly Ser Ala Leu Val Asp Asp Trp
                405                 410                 415
His Cys Leu Lys Thr Met Val Arg Thr Phe Glu Thr His Cys Gly Ser
            420                 425                 430
Leu Ser Gln Tyr Gly Met Lys His Met Arg Ser Phe Ala Asn Ile Cys
        435                 440                 445
Asn Val Gly Ile Lys Asn Glu Gln Met Ala Glu Ala Ser Ala Gln Ala
    450                 455                 460
Cys Val Ser Ile Pro Ser Asn Pro Trp Ser Ser Leu Gln Arg Gly Phe
465                 470                 475                 480
Ser Ala
```

<210> 11
<211> 1736
<212> DNA
<213> Glycine max

<220>
<221> CDS
<222> (41)...(1528)

<400> 11
gtgagtgacc gagtgagttt gtttttctca gctgatatat atg gcg ctt gat cgc 55
Met Ala Leu Asp Arg
1 5

tcc att ata agc aaa acg acg tgg tac agc gtc gta tta tgg atg atg 103
Ser Ile Ile Ser Lys Thr Thr Trp Tyr Ser Val Val Leu Trp Met Met
10 15 20

gtg gtg ctg gtg aga gtg cac ggt gca gcc gcg agg ccg aac cgg aag 151
Val Val Leu Val Arg Val His Gly Ala Ala Ala Arg Pro Asn Arg Lys
25 30 35

gag tgg gac tca gtc ata aag tta ccg act gaa ccg gtg gat gct gac 199
Glu Trp Asp Ser Val Ile Lys Leu Pro Thr Glu Pro Val Asp Ala Asp
40 45 50

tcg gat gaa gtg gga aca cga tgg gcg gtt ctc gtg gct ggt tca aac 247
Ser Asp Glu Val Gly Thr Arg Trp Ala Val Leu Val Ala Gly Ser Asn
55 60 65

ggc tac gga aac tac agg cat caa gca gat gtg tgc cat gcg tac cag 295
Gly Tyr Gly Asn Tyr Arg His Gln Ala Asp Val Cys His Ala Tyr Gln
70 75 80 85

ttg ctg ata aaa ggt gga cta aaa gaa gag aac ata gtg gtg ttt atg 343
Leu Leu Ile Lys Gly Gly Leu Lys Glu Glu Asn Ile Val Val Phe Met
90 95 100

tac gat gac ata gct acc aac gag ttg aat cct aga cat gga gtc atc 391
Tyr Asp Asp Ile Ala Thr Asn Glu Leu Asn Pro Arg His Gly Val Ile
105 110 115

atc aac cac cct gag gga gaa gat ctg tat gct ggt gtt cct aag gat 439
Ile Asn His Pro Glu Gly Glu Asp Leu Tyr Ala Gly Val Pro Lys Asp
120 125 130

tac acc ggt gat aat gtg acg acg gag aac ctc ttt gct gtt att ctt 487
Tyr Thr Gly Asp Asn Val Thr Thr Glu Asn Leu Phe Ala Val Ile Leu
135 140 145

gga gac aag agt aaa ttg aag gga gga agt ggc aaa gtg atc aac agc 535
Gly Asp Lys Ser Lys Leu Lys Gly Gly Ser Gly Lys Val Ile Asn Ser
150 155 160 165

aaa ccc gag gac aga ata ttt ata tac tac tct gat cat gga ggt cct 583
Lys Pro Glu Asp Arg Ile Phe Ile Tyr Tyr Ser Asp His Gly Gly Pro
170 175 180

gga ata ctt ggg atg cca aac atg cca tac ctt tat gcc atg gat ttt 631
Gly Ile Leu Gly Met Pro Asn Met Pro Tyr Leu Tyr Ala Met Asp Phe
185 190 195

att gat gtc ttg aag aag aaa cat gca tct gga agt tac aag gag atg 679
Ile Asp Val Leu Lys Lys Lys His Ala Ser Gly Ser Tyr Lys Glu Met
200 205 210

```
gtt ata tac gtg gaa gct tgt gaa agt ggg agc gtg ttt gag ggt ata   727
Val Ile Tyr Val Glu Ala Cys Glu Ser Gly Ser Val Phe Glu Gly Ile
    215                 220                 225

atg cct aag gat ctg aat att tat gtc aca act gca tca aat gca caa   775
Met Pro Lys Asp Leu Asn Ile Tyr Val Thr Thr Ala Ser Asn Ala Gln
230                 235                 240                 245

gag aat agt tgg ggg act tat tgt cct gga atg gat cct tct cca cct   823
Glu Asn Ser Trp Gly Thr Tyr Cys Pro Gly Met Asp Pro Ser Pro Pro
                250                 255                 260

cca gag tac atc act tgc cta ggg gat ttg tac agc gtt gct tgg atg   871
Pro Glu Tyr Ile Thr Cys Leu Gly Asp Leu Tyr Ser Val Ala Trp Met
            265                 270                 275

gaa gat agt gag gct cac aat cta aaa agg gaa tcc gtg aaa caa caa   919
Glu Asp Ser Glu Ala His Asn Leu Lys Arg Glu Ser Val Lys Gln Gln
        280                 285                 290

tac aaa tcg gta aag caa cgg act tca aat ttc aac aac tat gcg atg   967
Tyr Lys Ser Val Lys Gln Arg Thr Ser Asn Phe Asn Asn Tyr Ala Met
    295                 300                 305

ggt tct cat gtg atg caa tat ggt gat acc aac atc aca gct gaa aag   1015
Gly Ser His Val Met Gln Tyr Gly Asp Thr Asn Ile Thr Ala Glu Lys
310                 315                 320                 325

ctt tat tta tac caa ggt ttt gat cct gcc act gtg aac ttc cct cca   1063
Leu Tyr Leu Tyr Gln Gly Phe Asp Pro Ala Thr Val Asn Phe Pro Pro
                330                 335                 340

caa aac ggc agg cta gaa act aaa atg gaa gtt gtt aac caa aga gat   1111
Gln Asn Gly Arg Leu Glu Thr Lys Met Glu Val Val Asn Gln Arg Asp
            345                 350                 355

gca gaa ctt ttc tta ttg tgg caa atg tat cag aga tca aac cat cag   1159
Ala Glu Leu Phe Leu Leu Trp Gln Met Tyr Gln Arg Ser Asn His Gln
        360                 365                 370

tca gaa aat aag aca gac atc ctc aaa caa att gcg gag aca gtg aag   1207
Ser Glu Asn Lys Thr Asp Ile Leu Lys Gln Ile Ala Glu Thr Val Lys
    375                 380                 385

cat agg aaa cac ata gat ggt agc gtg gaa ttg att gga gtt tta ctg   1255
His Arg Lys His Ile Asp Gly Ser Val Glu Leu Ile Gly Val Leu Leu
390                 395                 400                 405

tat gga cca gga aaa ggt tct tct gtt cta caa tcc gtg agg gct cct   1303
Tyr Gly Pro Gly Lys Gly Ser Ser Val Leu Gln Ser Val Arg Ala Pro
                410                 415                 420

ggt tcg tcc ctt gtt gat gac tgg aca tgc cta aaa tca atg gtt cgg   1351
Gly Ser Ser Leu Val Asp Asp Trp Thr Cys Leu Lys Ser Met Val Arg
            425                 430                 435

gtg ttt gaa act cac tgt ggg aca ctg act cag tat ggc atg aaa cac   1399
```

```
Val Phe Glu Thr His Cys Gly Thr Leu Thr Gln Tyr Gly Met Lys His
        440                 445                 450

atg cga gca ttc gcc aac att tgc aac agt ggc gtt tct gag gcc tcc   1447
Met Arg Ala Phe Ala Asn Ile Cys Asn Ser Gly Val Ser Glu Ala Ser
    455                 460                 465

atg gaa gag gct tgt ttg gca gcc tgt gaa ggc tac aat gct ggg cta   1495
Met Glu Glu Ala Cys Leu Ala Ala Cys Glu Gly Tyr Asn Ala Gly Leu
470                 475                 480                 485

ttc cat cca tca aac aga ggc tac agt gct tga ttttgggttt tgtacacaaa 1548
Phe His Pro Ser Asn Arg Gly Tyr Ser Ala  *
                490                 495

agctttaaag cccggttgat gatgtaatat ttctctattg cattctgcct actggtttct 1608
gctgcttgtg tcaaattttc tctaaactag agtagcccaa tagcatacgt gttatgtgca 1668
ttggtcatgt atacaagtgt aatactaata ccttcctaca taatataaga ttagttagtt 1728
tacttgtc                                                          1736

<210> 12
<211> 495
<212> PRT
<213> Glycine max

<400> 12
Met Ala Leu Asp Arg Ser Ile Ile Ser Lys Thr Thr Trp Tyr Ser Val
 1               5                  10                  15
Val Leu Trp Met Met Val Val Leu Val Arg Val His Gly Ala Ala Ala
            20                  25                  30
Arg Pro Asn Arg Lys Glu Trp Asp Ser Val Ile Lys Leu Pro Thr Glu
        35                  40                  45
Pro Val Asp Ala Asp Ser Asp Glu Val Gly Thr Arg Trp Ala Val Leu
    50                  55                  60
Val Ala Gly Ser Asn Gly Tyr Gly Asn Tyr Arg His Gln Ala Asp Val
65                  70                  75                  80
Cys His Ala Tyr Gln Leu Leu Ile Lys Gly Gly Leu Lys Glu Glu Asn
                85                  90                  95
Ile Val Val Phe Met Tyr Asp Asp Ile Ala Thr Asn Glu Leu Asn Pro
            100                 105                 110
Arg His Gly Val Ile Ile Asn His Pro Glu Gly Glu Asp Leu Tyr Ala
        115                 120                 125
Gly Val Pro Lys Asp Tyr Thr Gly Asp Asn Val Thr Thr Glu Asn Leu
    130                 135                 140
Phe Ala Val Ile Leu Gly Asp Lys Ser Lys Leu Lys Gly Gly Ser Gly
145                 150                 155                 160
Lys Val Ile Asn Ser Lys Pro Glu Asp Arg Ile Phe Ile Tyr Tyr Ser
                165                 170                 175
Asp His Gly Gly Pro Gly Ile Leu Gly Met Pro Asn Met Pro Tyr Leu
            180                 185                 190
Tyr Ala Met Asp Phe Ile Asp Val Leu Lys Lys Lys His Ala Ser Gly
        195                 200                 205
Ser Tyr Lys Glu Met Val Ile Tyr Val Glu Ala Cys Glu Ser Gly Ser
    210                 215                 220
Val Phe Glu Gly Ile Met Pro Lys Asp Leu Asn Ile Tyr Val Thr Thr
225                 230                 235                 240
Ala Ser Asn Ala Gln Glu Asn Ser Trp Gly Thr Tyr Cys Pro Gly Met
```

```
                245                 250                 255
Asp Pro Ser Pro Pro Pro Glu Tyr Ile Thr Cys Leu Gly Asp Leu Tyr
            260                 265                 270
Ser Val Ala Trp Met Glu Asp Ser Glu Ala His Asn Leu Lys Arg Glu
        275                 280                 285
Ser Val Lys Gln Gln Tyr Lys Ser Val Lys Gln Arg Thr Ser Asn Phe
    290                 295                 300
Asn Asn Tyr Ala Met Gly Ser His Val Met Gln Tyr Gly Asp Thr Asn
305                 310                 315                 320
Ile Thr Ala Glu Lys Leu Tyr Leu Tyr Gln Gly Phe Asp Pro Ala Thr
                325                 330                 335
Val Asn Phe Pro Pro Gln Asn Gly Arg Leu Glu Thr Lys Met Glu Val
            340                 345                 350
Val Asn Gln Arg Asp Ala Glu Leu Phe Leu Leu Trp Gln Met Tyr Gln
        355                 360                 365
Arg Ser Asn His Gln Ser Glu Asn Lys Thr Asp Ile Leu Lys Gln Ile
    370                 375                 380
Ala Glu Thr Val Lys His Arg Lys His Ile Asp Gly Ser Val Glu Leu
385                 390                 395                 400
Ile Gly Val Leu Leu Tyr Gly Pro Gly Lys Gly Ser Ser Val Leu Gln
                405                 410                 415
Ser Val Arg Ala Pro Gly Ser Ser Leu Val Asp Asp Trp Thr Cys Leu
            420                 425                 430
Lys Ser Met Val Arg Val Phe Glu Thr His Cys Gly Thr Leu Thr Gln
        435                 440                 445
Tyr Gly Met Lys His Met Arg Ala Phe Ala Asn Ile Cys Asn Ser Gly
    450                 455                 460
Val Ser Glu Ala Ser Met Glu Glu Ala Cys Leu Ala Ala Cys Glu Gly
465                 470                 475                 480
Tyr Asn Ala Gly Leu Phe His Pro Ser Asn Arg Gly Tyr Ser Ala
                485                 490                 495
```

<210> 13
<211> 1715
<212> DNA
<213> Glycine max

<220>
<221> CDS
<222> (19)...(1509)

<400> 13

```
tgttgctgtc gagctgat atg gcg gtt gat cgc tcc ctt acg agg tgc tgt   51
                      Met Ala Val Asp Arg Ser Leu Thr Arg Cys Cys
                       1               5                  10

agc ctc gta ctg tgg tcg tgg atg ttg ctg agg atg atg atg gcg cag   99
Ser Leu Val Leu Trp Ser Trp Met Leu Leu Arg Met Met Met Ala Gln
             15                  20                  25

ggt gca gcc gcg agg gcc aac cgg aag gag tgg gac tcg gtc ata aag   147
Gly Ala Ala Ala Arg Ala Asn Arg Lys Glu Trp Asp Ser Val Ile Lys
         30                  35                  40

tta ccg gct gaa ccg gtc gat gct gac tcg gat cat gaa gtg gga aca   195
Leu Pro Ala Glu Pro Val Asp Ala Asp Ser Asp His Glu Val Gly Thr
```

```
     45                  50                  55

cga tgg gcg gtt ctt gtg gct ggt tca aac ggc tat gga aac tac agg   243
Arg Trp Ala Val Leu Val Ala Gly Ser Asn Gly Tyr Gly Asn Tyr Arg
 60                  65                  70                  75

cat caa gca gat gtg tgc cat gcg tac cag ttg ctg ata aaa ggt ggg   291
His Gln Ala Asp Val Cys His Ala Tyr Gln Leu Leu Ile Lys Gly Gly
                 80                  85                  90

cta aaa gaa gag aac ata gtg gtg ttt atg tac gat gac ata gct aca   339
Leu Lys Glu Glu Asn Ile Val Val Phe Met Tyr Asp Asp Ile Ala Thr
             95                 100                 105

gac gag tta aat ccc aga cct gga gtc atc atc aac cac cct gag gga   387
Asp Glu Leu Asn Pro Arg Pro Gly Val Ile Ile Asn His Pro Glu Gly
        110                 115                 120

caa gat gtg tat gct ggt gtt cct aag gat tac acc ggt gag aat gtg   435
Gln Asp Val Tyr Ala Gly Val Pro Lys Asp Tyr Thr Gly Glu Asn Val
    125                 130                 135

acg gcc cag aac ctc ttt gcc gtt att ctt gga gac aag aat aaa gtg   483
Thr Ala Gln Asn Leu Phe Ala Val Ile Leu Gly Asp Lys Asn Lys Val
140                 145                 150                 155

aag gga gga agt ggc aaa gtg atc aat agc aaa cct gag gac aga ata   531
Lys Gly Gly Ser Gly Lys Val Ile Asn Ser Lys Pro Glu Asp Arg Ile
                160                 165                 170

ttt ata tac tac tct gat cat gga ggt ccg gga gtt ctt ggg atg cca   579
Phe Ile Tyr Tyr Ser Asp His Gly Gly Pro Gly Val Leu Gly Met Pro
            175                 180                 185

aac atg cca tac ctt tat gct atg gac ttt att gaa gtc ttg aag aag   627
Asn Met Pro Tyr Leu Tyr Ala Met Asp Phe Ile Glu Val Leu Lys Lys
        190                 195                 200

aaa cat gca tct gga ggt tac aag aag atg gtc ata tac gtg gaa gct   675
Lys His Ala Ser Gly Gly Tyr Lys Lys Met Val Ile Tyr Val Glu Ala
    205                 210                 215

tgt gaa agt ggg aac cat gtt ttg aag ggt ata atg cct aag gat ctg   723
Cys Glu Ser Gly Asn His Val Leu Lys Gly Ile Met Pro Lys Asp Leu
220                 225                 230                 235

cag att tat gtc aca act gca tca aat gca caa gag aat agt tgg gga   771
Gln Ile Tyr Val Thr Thr Ala Ser Asn Ala Gln Glu Asn Ser Trp Gly
                240                 245                 250

act tat tgt cct gga atg gat cct tct cca cct cca gag tac atc act   819
Thr Tyr Cys Pro Gly Met Asp Pro Ser Pro Pro Pro Glu Tyr Ile Thr
            255                 260                 265

tgc cta ggg gat ttg tac agt gtt gct tgg atg gaa gat agt gag act   867
Cys Leu Gly Asp Leu Tyr Ser Val Ala Trp Met Glu Asp Ser Glu Thr
        270                 275                 280
```

```
cat aat cta aaa agg gag tcc gtg aaa caa caa tac aaa tcg gta aag   915
His Asn Leu Lys Arg Glu Ser Val Lys Gln Gln Tyr Lys Ser Val Lys
    285                 290                 295

caa cgg act tca aat ttc aac aac tat gcg atg ggt tct cat gtg atg   963
Gln Arg Thr Ser Asn Phe Asn Asn Tyr Ala Met Gly Ser His Val Met
300                 305                 310                 315

caa tac ggt gac aca aac atc aca gct gaa aag ctt tat tta tac caa   1011
Gln Tyr Gly Asp Thr Asn Ile Thr Ala Glu Lys Leu Tyr Leu Tyr Gln
                320                 325                 330

ggt ttt gat cct gcc gct gtg aac ttc cct cca cag aac gga agg cta   1059
Gly Phe Asp Pro Ala Ala Val Asn Phe Pro Pro Gln Asn Gly Arg Leu
            335                 340                 345

gaa act aaa atg gaa gtt gtt aac caa aga gat gca gaa ctt ttc ttc   1107
Glu Thr Lys Met Glu Val Val Asn Gln Arg Asp Ala Glu Leu Phe Phe
        350                 355                 360

atg tgg caa atg tat cag aga tca aac cat cag cca gaa aag aag aca   1155
Met Trp Gln Met Tyr Gln Arg Ser Asn His Gln Pro Glu Lys Lys Thr
    365                 370                 375

gac atc ctc aaa cag ata gcg gag aca gtg aag cat agg aaa cac ata   1203
Asp Ile Leu Lys Gln Ile Ala Glu Thr Val Lys His Arg Lys His Ile
380                 385                 390                 395

gat ggt agc gtg gaa ttg att gga gtt tta ttg tat gga cca gga aaa   1251
Asp Gly Ser Val Glu Leu Ile Gly Val Leu Leu Tyr Gly Pro Gly Lys
                400                 405                 410

ggt tct tct gtt cta caa tcc atg agg gct cct ggt ctt gcc ctt gtt   1299
Gly Ser Ser Val Leu Gln Ser Met Arg Ala Pro Gly Leu Ala Leu Val
            415                 420                 425

gat gac tgg aca tgc cta aaa tca atg gtt cgg gtg ttt gag act cac   1347
Asp Asp Trp Thr Cys Leu Lys Ser Met Val Arg Val Phe Glu Thr His
        430                 435                 440

tgt ggg aca ctg act cag tat ggc atg aaa cac atg cga gca ttt gcc   1395
Cys Gly Thr Leu Thr Gln Tyr Gly Met Lys His Met Arg Ala Phe Ala
    445                 450                 455

aac att tgc aac agc ggt gtt tct gag gcc tcc atg gaa gag gtt tgt   1443
Asn Ile Cys Asn Ser Gly Val Ser Glu Ala Ser Met Glu Glu Val Cys
460                 465                 470                 475

gtg gca gct tgt gaa ggc tac gat tct ggg cta tta cat cca tca aac   1491
Val Ala Ala Cys Glu Gly Tyr Asp Ser Gly Leu Leu His Pro Ser Asn
                480                 485                 490

aaa ggc tat agt gct tga ttttgggttt tgtacacagc ttaaaaaccc          1539
Lys Gly Tyr Ser Ala  *
            495
```

ggttgatgat gtaatacttc tctattgcat tctccctact ggtttctgct gcatgtgtca 1599
aattttctct aaactagagt agcccaatag catacgtgtt atgagcattg gtcatgtata 1659
taagtgtaat agtaatatct tttacatatt ataagatcag ttagtttggt ttacta 1715

<210> 14
<211> 496
<212> PRT
<213> Glycine max

<400> 14
Met Ala Val Asp Arg Ser Leu Thr Arg Cys Cys Ser Leu Val Leu Trp
1 5 10 15
Ser Trp Met Leu Leu Arg Met Met Met Ala Gln Gly Ala Ala Ala Arg
20 25 30
Ala Asn Arg Lys Glu Trp Asp Ser Val Ile Lys Leu Pro Ala Glu Pro
35 40 45
Val Asp Ala Asp Ser Asp His Glu Val Gly Thr Arg Trp Ala Val Leu
50 55 60
Val Ala Gly Ser Asn Gly Tyr Gly Asn Tyr Arg His Gln Ala Asp Val
65 70 75 80
Cys His Ala Tyr Gln Leu Leu Ile Lys Gly Gly Leu Lys Glu Glu Asn
85 90 95
Ile Val Val Phe Met Tyr Asp Asp Ile Ala Thr Asp Glu Leu Asn Pro
100 105 110
Arg Pro Gly Val Ile Ile Asn His Pro Glu Gly Gln Asp Val Tyr Ala
115 120 125
Gly Val Pro Lys Asp Tyr Thr Gly Glu Asn Val Thr Ala Gln Asn Leu
130 135 140
Phe Ala Val Ile Leu Gly Asp Lys Asn Lys Val Lys Gly Gly Ser Gly
145 150 155 160
Lys Val Ile Asn Ser Lys Pro Glu Asp Arg Ile Phe Ile Tyr Tyr Ser
165 170 175
Asp His Gly Gly Pro Gly Val Leu Gly Met Pro Asn Met Pro Tyr Leu
180 185 190
Tyr Ala Met Asp Phe Ile Glu Val Leu Lys Lys Lys His Ala Ser Gly
195 200 205
Gly Tyr Lys Lys Met Val Ile Tyr Val Glu Ala Cys Glu Ser Gly Asn
210 215 220
His Val Leu Lys Gly Ile Met Pro Lys Asp Leu Gln Ile Tyr Val Thr
225 230 235 240
Thr Ala Ser Asn Ala Gln Glu Asn Ser Trp Gly Thr Tyr Cys Pro Gly
245 250 255
Met Asp Pro Ser Pro Pro Pro Glu Tyr Ile Thr Cys Leu Gly Asp Leu
260 265 270
Tyr Ser Val Ala Trp Met Glu Asp Ser Glu Thr His Asn Leu Lys Arg
275 280 285
Glu Ser Val Lys Gln Gln Tyr Lys Ser Val Lys Gln Arg Thr Ser Asn
290 295 300
Phe Asn Asn Tyr Ala Met Gly Ser His Val Met Gln Tyr Gly Asp Thr
305 310 315 320
Asn Ile Thr Ala Glu Lys Leu Tyr Leu Tyr Gln Gly Phe Asp Pro Ala
325 330 335
Ala Val Asn Phe Pro Pro Gln Asn Gly Arg Leu Glu Thr Lys Met Glu
340 345 350
Val Val Asn Gln Arg Asp Ala Glu Leu Phe Phe Met Trp Gln Met Tyr
355 360 365
Gln Arg Ser Asn His Gln Pro Glu Lys Lys Thr Asp Ile Leu Lys Gln

```
    370                 375                 380
Ile Ala Glu Thr Val Lys His Arg Lys His Ile Asp Gly Ser Val Glu
385                 390                 395                 400
Leu Ile Gly Val Leu Leu Tyr Gly Pro Gly Lys Gly Ser Ser Val Leu
                405                 410                 415
Gln Ser Met Arg Ala Pro Gly Leu Ala Leu Val Asp Asp Trp Thr Cys
            420                 425                 430
Leu Lys Ser Met Val Arg Val Phe Glu Thr His Cys Gly Thr Leu Thr
        435                 440                 445
Gln Tyr Gly Met Lys His Met Arg Ala Phe Ala Asn Ile Cys Asn Ser
    450                 455                 460
Gly Val Ser Glu Ala Ser Met Glu Glu Val Cys Val Ala Ala Cys Glu
465                 470                 475                 480
Gly Tyr Asp Ser Gly Leu Leu His Pro Ser Asn Lys Gly Tyr Ser Ala
                485                 490                 495
```

<210> 15
<211> 7825
<212> DNA
<213> Artificial Sequence

<220>
<223> Expression cassette for suppression of soybean VPE expression

<221> misc_feature
<222> 1223, 6120, 6366
<223> N=A, T, C, or G

<221> promoter
<222> (53)...(2138)
<223> Promoter from Kuntz trypsin inhibitor gene of Glycine Max

<221> misc_feature
<222> (2141)...(2170)
<223> EL (stem) forward sequence

<221> misc_feature
<222> (2192)...(2761)
<223> VPE2A fragment

<221> misc_feature
<222> (2762)...(3273)
<223> VPE2B fragment

<221> misc_feature
<222> (3274)...(3566)
<223> VPE3A fragment

<221> misc_feature
<222> (3584)...(3884)
<223> VPE1A fragment

<221> misc_feature
<222> (3891)...(4290)

<223> VPE1B fragment

<221> misc_feature
<222> (4290)...(4319)
<223> EL (stem) reverse sequence

<221> terminator
<222> (4343)...(4545)
<223> Terminator from Kuntz trypsin inhibitor gene of
Glycine max

<221> CDS
<222> (5815)...(6630)
<223> Kanamycin resistance gene

<400> 15
gggcgaattg ggttacccgg accggaattc gagctcggta cccggggatc ctcgaagaga 60
agggttaata acacattttt taacattttt aacacaaatt ttagttattt aaaaatttat 120
taaaaaattt aaaataagaa gaggaactct ttaaataaat ctaacttaca aaatttatga 180
tttttaataa gttttcacca ataaaaaatg tcataaaaat atgttaaaaa gtatattatc 240
aatattctct ttatgataaa taaaaagaaa aaaaaaataa aagttaagtg aaaatgagat 300
tgaagtgact ttaggtgtgt ataaatatat caaccccgcc aacaatttat ttaatccaaa 360
tatattgaag tatattattc catagccttt atttatttat atatttatta tataaaagct 420
ttatttgttc taggttgttc atgaaatatt tttttggttt tatctccgtt gtaagaaaat 480
catgtgcttt gtgtcgccac tcactattgc agctttttca tgcattggtc agattgacgg 540
ttgattgtat ttttgttttt tatggttttg tgttatgact taagtcttca tctctttatc 600
tcttcatcag gtttgatggt tacctaatat ggtccatggg tacatgcatg gttaaattag 660
gtggccaact ttgttgtgaa cgatagaatt ttttttatat taagtaaact atttttatat 720
tatgaaataa taataaaaaa aatattttat cattattaac aaaatcatat tagttaattt 780
gttaactcta taataaaaga aatactgtaa cattcacatt acatggtaac atctttccac 840
cctttcattt gttttttgtt tgatgacttt ttttcttgtt taaatttatt tcccttcttt 900
taaatttgga atacattatc atcatatata aactaaaata ctaaaaacag gattacacaa 960
atgataaata ataacacaaa tatttataaa tctagctgca atatatttaa actagctata 1020
tcgatattgt aaaataaaac tagctgcatt gatactgata aaaaaatatc atgtgctttc 1080
tggactgatg atgcagtata cttttgacat tgcctttatt ttatttttca gaaaagcttt 1140
cttagttctg ggttcttcat tatttgtttc ccatctccat tgtgaattga atcatttgct 1200
tcgtgtcaca aatacaattt agntaggtac atgcattggt cagattcacg gtttattatg 1260
tcatgactta agttcatggt agtacattac ctgccacgca tgcattatat tggttagatt 1320
tgataggcaa atttggttgt caacaatata aatataaata atgtttttat attacgaaat 1380
aacagtgatc aaaacaaaca gttttatctt tattaacaag attttgtttt tgtttgatga 1440
cgtttttttaa tgtttacgct ttccccccttc ttttgaattt agaacacttt atcatcataa 1500
aatcaaatac taaaaaaatt acatatttca taaataataa cacaaatatt tttaaaaaat 1560
ctgaaataat aatgaacaat attacatatt atcacgaaaa ttcattaata aaaatattat 1620
ataaataaaa tgtaatagta gttatatgta ggaaaaaagt actgcacgca taatatatac 1680
aaaaagatta aaatgaacta ttataaataa taacactaaa ttaatggtga atcatatcaa 1740
aataatgaaa aagtaaataa aatttgtaat taacttctat atgtattaca cacacaaata 1800
ataaataata gtaaaaaaaa ttatgataaa tatttaccat ctcataagat atttaaaata 1860
atgataaaaa tatagattat tttttatgca actagctagc caaaaagaga acacgggtat 1920
atataaaaag agtaccttta aattctactg tacttccttt attcctgacg tttttatatc 1980
aagtggacat acgtgaagat tttaattatc agtctaaata tttcattagc acttaatact 2040
tttctgtttt attcctatcc tataagtagt cccgattctc ccaacattgc ttattcacac 2100
aactaactaa gaaagtcttc catagccccc caagcggccg gagctggtca tctcgctcat 2160
cgtcgagtcg gcggccgctc tagaactagt ggatcccccg ggctgcagga attcgatgca 2220
cgagaattaa attaatagag gatgaaattc tagtttaagg aaggttggtt ggttgggtgg 2280
gggtaggaga tactctcatt cacctcccat catcattata atcattcatt ccaacctacc 2340
cttattcttc ttcttcaatt tcacacccat catggaccgt tttccgatcc tctttctcgt 2400
cgccaccctc atcaccctcg cctccggtgc ccgccacgat attctccggt taccctccga 2460

agcttccagg ttcttcaaag cacctgctaa tgccgatcaa aacgatgagg gcaccaggtg 2520
ggccgtttta gttgccggtt ccaatggcta ctggaattac aggcaccagt ctgatgtttg 2580
ccatgcatat caactactga ggaaaggtgg tgtgaaagag gaaaatattg ttgtatttat 2640
gtatgatgac attgctttca atgaagagaa cccacggcct ggagtcatta ttaacagtcc 2700
acacggaaat gatgtttaca agggagttcc taaggattac gttggtgaag atgttactgt 2760
taaccaacgt gatgcagatc tcatccattt ttgggataag ttccgcaaag ctcctgtggg 2820
ttcttctagg aaagctgcag ctgagaaaca aattcttgaa gcaatgtctc acagaatgca 2880
tatagatgac agcatgaaac gtattggaaa gctcttcttt ggcattgaaa agggtccaga 2940
actgcttagc agtgttagac ctgctgggca accacttgtt gatgactggg actgccttaa 3000
aacattggtt aggacttttg agacacattg tggatccctg tctcagtatg ggatgaaaca 3060
tatgaggtcc tttgcaaact tctgcaacgc tggaatacga aaagagcaaa tggctgaggc 3120
ctcagcacaa gcatgtgtca atatccctgc tagttcctgg agttctatgc acaggggttt 3180
cagtgcataa ttcctagaat gcgctccatt gaagaccgag tatagtcgtt gtaacattat 3240
tctttacgag tgttatggac tgtactctct gctcatggtg aggacttttg agacacattg 3300
tggatccttg tctcaatacg ggatgaaaca catgaggtcc tttgcaaaca tctgcaatgt 3360
agggataaag aatgaacaaa tggctgaggc ttcagcacaa gcttgtgtca gtattccttc 3420
caatccctgg agttctctgc aaaggggttt cagtgcataa taactccctg taatgtgcac 3480
tagtaaagac caaagtatga ttattgttac attatgttac atggttgtac ttgtatatac 3540
atatcttgtc ccacctttgt aaatacaatt cgatgggctg caggaattcg atgtgagtga 3600
ccgagtgagt ttgtttttct cagctgatat atatggcgct tgatcgctcc attataagca 3660
aaacgacgtg gtacagcgtc gtattatgga tgatggtggt gctggtgaga gtgcacggtg 3720
cagccgcgag gccgaaccgg aaggagtggg actcagtcat aaagttaccg actgaaccgg 3780
tggatgctga ctcggatgaa gtgggaacac gatgggcggt tctcgtggct ggttcaaacg 3840
gctacggaaa ctacaggcat caagcagatg tgtgccatgc gtaccagttg ctgccacgag 3900
gtgagtcttt cttagctgat atggcggttg atcgctccct tacgaggtgc tgtagcctcg 3960
tactgtggtc gtggatgttg ctgaggatga tgatggcgca gggtgcagcc gcgagggcca 4020
accggaagga gtgggactcc atggaagagg tttgtgtggc agcttgtgaa ggctacgatt 4080
ctgggctatt acatccatca aacaaaggct atagtgcttg attttgggtt ttgtacacag 4140
cttaaaaacc cggttgatga tgtaatactt ctctattgca ttctccctac tggtttctgc 4200
tgcatgtgtc aaattttctc taaactagag tagcccaata gcatacgtgt tatgagcatt 4260
ggtcatgtat ataagtgtaa tagtaatatc gactcgacga tgagcgagat gaccagctcg 4320
aattcatcac tagtgaattc gcggccgcga cacaagtgtg agagtactaa ataaatgctt 4380
tggttgtacg aaatcattac actaaataaa ataatcaaag cttatatatg ccttccgcta 4440
aggccgaatg caaagaaatt ggttctttct cgttatcttt tgccactttt actagtacgt 4500
attaattact acttaatcat ctttgtttac ggctcattat atccgtcgac ctcgaggggg 4560
ggccccggcc gaagcttcgg tccgggtcac ccagcttgag tattctatag tgtcacctaa 4620
atagcttggc gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa 4680
ttccacacaa catacgagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga 4740
gctaactcac attaattgcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt 4800
gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgct 4860
cttccgcttc ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat 4920
cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac gcaggaaaga 4980
acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt 5040
ttttcgatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt 5100
ggcgaaaccc gacaggacta taaagatacc aggcgtttcc ccctggaagc tccctcgtgc 5160
gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa 5220
gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct 5280
ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc ttatccggta 5340
actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg 5400
gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc 5460
ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta 5520
ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg 5580
gttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt 5640
tgatcttttc tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg 5700
tcatggagcc acgttgtgtc tcaaaatctc tgatgttaca ttgcacaaga taaaaatata 5760
tcatcatgaa caataaaact gtctgcttac ataaacagta atacaagggg tgttatgagc 5820
catattcaac gggaaacgtc ttgctcgagg ccgcgattaa attccaacat ggatgctgat 5880

ttatatgcct ataaatgggc tcgcgataat gtcggccaat caggtccgac aatctatcga 5940
ttgtatggga agcccgatgc gccagacttg tttctgaaac atggcaaagg tagccttgcc 6000
aatgatgtta cagatgagat ggtcagacta aactgcctga cggaatttat gcctcttccg 6060
accatcaagc attttatccg tactcctgat gatgcatggt tactcaccac tgcgatcccn 6120
gggaaaacag cattccaggt attagaagaa tatcctgatt caggtgaaaa tattgttgat 6180
gcgctggcag tgttcctgcg ccggttgcat tcgattcctc tttgtaattg tccttttaac 6240
agcgatcgcg tatttcgtct cgctcaggcg caatcacgaa tgaataacgg tttggttgat 6300
gcgagtgatt ttgatgacga gcgtaatggc tggcctgttg aacaagtctg gaaagaaatg 6360
cataancttt tgccattctc accggattca gtcgtcactc atggtgattt ctcacttgat 6420
aaccttattt ttgaccaggc gaaattaata ggttgtattg atcttcgacg agtcggaatc 6480
gcagaccgat accaggatct tgccatccta tggaactgcc tcggtgagtt ttctccttca 6540
ttacagaaac ggctttttca aaaatatggt attgataatc ctgatatgaa taaattgcag 6600
tttcatttga tcctcgatga gtttttctaa tcagaattgg ttaattggtt gtaacactgg 6660
cagagcatta cgctgacttg acgggacggc ggctttgttg aataaatcga acttttgctg 6720
acttgaagga tcagatcacg catcttcccg acaacgcaga ccgttccgtg gcaaagcaaa 6780
agttcaaaat caccaactgg tccacctaca acaaagctct catcaaccgt ggctccctca 6840
ctttctggct ggatgatggg gcgattcagg cctggtatga gtcagcaaca ccttcttcac 6900
gagccatgac attaacctat aaaaataggc gtatcacgag gccctttcgt ctcgcgcgtt 6960
tcggtgatga cggtgaaaac ctctgacaca tgcagctccc ggagacggtc acagcttgtc 7020
tgtaagcgga tgccgggagc agacaagccc gtcagggcgc gtcagcgggt gttggcgggt 7080
gtcggggctg gcttaactat gcggcatcag agcagattgt actgagagtg caccatatgc 7140
ggtgtgaaat accgcacaga tgcgtaagga gaaaataccg catcaggcga attgtaaac 7200
gttaatattt tgttaaaatt cgcgttaaat atttgttaaa tcagctcatt ttttaaccaa 7260
taggccgaaa tcggcaaaat cccttataaa tcaaaagaat agaccgagat agggttgagt 7320
gttgttccag tttggaacaa gagtccacta ttaaagaacg tggactccaa cgtcaaaggg 7380
cgaaaaaccg tctatcaggg cgatggccca ctacgtgaac catcacccaa atcaagtttt 7440
ttgcggtcga ggtgccgtaa agctctaaat cggaacccta aagggagccc ccgatttaga 7500
gcttgacggg gaaagccggc gaacgtggcg agaaaggaag ggaagaaagc gaaaggagcg 7560
ggcgctaggg cgctggcaag tgtagcggtc acgctgcgcg taaccaccac acccgccgcg 7620
cttaatgcgc cgctacaggg cgcgtccatt cgccattcag gctgcgcaac tgttgggaag 7680
ggcgatcggt gcgggcctct tcgctattac gccagctggc gaaaggggga tgtgctgcaa 7740
ggcgattaag ttgggtaacg ccagggtttt cccagtcacg acgttgtaaa acgacggcca 7800
gtgaattgta atacgactca ctata 7825

METHODS AND COMPOSITIONS FOR ALTERING THE FUNCTIONAL PROPERTIES OF SEED STORAGE PROTEINS IN SOYBEAN

FIELD OF THE INVENTION

The present invention relates to genetic modification of soybean, more particularly to the alteration of the functional properties seed storage proteins in soybean.

BACKGROUND OF THE INVENTION

Many plant storage tissues (seeds, leaves, roots, and tubers), accumulate sizable reserves of proteins during development. For example, cultivated soybean seeds contain an average of about 40% protein, and in some varieties protein levels reach as much as 55% of the dry weight. The abundance of proteins in legume seeds has made them the primary dietary protein source and has stimulated an interest in developing approaches to genetically engineer seeds to improve their nutritional quality.

Plant storage proteins, especially those processed through the secretory pathway, generally undergo multiple post-translational processing steps including folding, assembly, intracellular sorting, and proteolytic processing, prior to final deposition (Müntz *et al.*, (1993) *Proc. Phytochem. Soc. Eur.* 35:128-146; Müntz (1998) *Plant Mol. Biol.* 38:77-99; Herman and Larkins (1999) *Plant Cell* 11:601-613). Accumulation and deposition of the proteins is accomplished by compartmentalization in specialized vacuoles termed protein storage vacuoles and or protein bodies (Hara-Nishimura *et al.* (1995) *J. Plant Physiol.* 145:632-640; Müntz (1998) *Plant Molec. Biol.* 38:77-99; Herman and Larkins (1999) *Plant Cell* 11:601-613).

The proteolytic processing steps of protein deposition in vacuoles include specific polypeptide cleavage steps accomplished by proteases localized to the storage vacuole (Bassham *et al.* (2000) *Curr. Opin. Cell Biol.* 12:491-495). Storage proteins that accumulate in vacuoles have therefore co-evolved with the environment of the storage vacuole, such that only a select few protease sites exist or are accessible to these proteases (Hara-Nishimura *et al.* (1987) *Plant Physiol.* 85:440-445; D'Hondt *et al.*, (1993) *J. Biol. Chem.* 268:10884-10891; Hara-Nishimura *et al.* (1993) *Plant Cell* 5:1651-1659; Hara-Nishimura *et al.* (1995) *J. Plant Physiol.* 145:632-640).

Glycinin is a major soybean seed storage protein that is used extensively in soy food products. However, this protein's functional properties limit its use in some product applications. For example, glycinin is insoluble at low pH, and so it is not well suited for use in acidic food products. See, for example, Lakemond *et al.* (2000) *J. Agric. Food Chem.* 48:1985-90 and Mohamed *et al.* (2002) *J. Agric. Food Chem.* 50:7380–85.

Accordingly, methods are needed to alter the functional properties of seed storage proteins in soybeans.

SUMMARY OF THE INVENTION

The present invention is directed to altering the functional properties of soybean seed storage proteins. It is the novel finding of the present invention that the functional properties of seed storage proteins can be altered by reducing the expression of one or more vacuolar processing enzymes (VPEs) in plant seed. Accordingly, in one embodiment, the invention provides a plant that is genetically modified to alter one or more functional properties of one or more seed storage proteins. The invention also provides methods for altering the functional properties of one or more soybean seed storage proteins. In some embodiments, the method comprises transforming a soybean plant cell with at least one expression cassette capable of expressing a polynucleotide that reduces or eliminates the activity of a vacuolar processing enzyme in the seed of said soybean plant, regenerating a transformed plant from the transformed plant cell, and collecting seed from the regenerated transformed plant. In other embodiments, the method comprises transforming a soybean plant cell with at least one expression cassette

comprising a polynucleotide encoding a polypeptide that reduces or eliminates the activity of at least one vacuolar processing enzyme in seed in the seed of said soybean plant, regenerating a transformed plant from the transformed plant cell, and collecting seed from the regenerated transformed plant.

According to the invention, the activity of at least one, at least two, at least three, at least four, at least five, or at least six vacuolar processing enzymes may be reduced or eliminated in soybean seed. Thus, the soybean plants may be transformed with two or more polynucleotides, which inhibit the expression of a soybean vacuolar processing enzyme. In some embodiments, the polynucleotide is designed to reduce or eliminate the activity of only one vacuolar processing enzyme, while in other embodiments the polynucleotide is designed to reduce or eliminate the expression of two or more different soybean vacuolar processing enzymes, three or more different soybean vacuolar processing enzymes, or more than three different soybean vacuolar processing enzymes. When two or more polynucleotides are transformed into the same plant cell, they may be expressed from the same expression cassette. Alternatively, the polynucleotides may be comprised in separate expression cassettes.

In some embodiments, at least one of the soybean vacuolar processing enzymes whose activity is reduced or eliminated is selected from the group consisting of soybean Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a, and Vpe3b. In further embodiments, at least one vacuolar processing enzyme whose expression is inhibited is selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14.

In certain embodiments, at least one functional property that is altered in the seed storage protein is the solubility of the seed storage protein. In particular embodiments, the solubility of a seed storage protein is increased at low pH. For example, the invention provides embodiments in which the solubility of the seed storage protein is increased between pH 4.0 and pH 6.0

In some embodiments, the soybean seed storage protein whose functional properties are altered is selected from glycinin, soybean 2S albumin, and β-conglycinin.

The expression cassettes used in the method of invention may be any expression cassette capable of reducing or eliminating the expression of at least one soybean vacuolar processing enzyme.

The invention also provides soybean plants that are genetically modified to alter the functional properties of one or more seed storage proteins. In some embodiments, the soybean plant is genetically modified to reduce or eliminate the expression of one or more vacuolar processing enzymes in seed. In particular embodiments, the soybean plant is stably transformed with an expression cassette capable of expressing at least one polynucleotide that inhibits the expression of a vacuolar processing enzyme in seed. In other embodiments, the soybean plant is stably transformed with at least one polynucleotide comprising a polynucleotide encoding a polypeptide that inhibits the activity of a vacuolar processing enzyme.

The soybean plant of the invention may be genetically modified to reduce or eliminate the activity of at least one, at least two, at least three, at least four, at least five, at least six, or at least seven or more soybean vacuolar processing enzymes. Transgenic seed of the genetically modified plant is also encompassed.

EMBODIMENTS OF THE INVENTION INCLUDE:

1. A soybean plant that is genetically modified to alter one or more functional properties of one or more seed storage proteins, wherein said soybean plant is genetically modified to reduce or eliminate the activity of one or more vacuolar processing enzymes in its seed.

2. The plant of 1, wherein said soybean plant is stably transformed with at least one expression cassette capable of expressing a polynucleotide that inhibits the expression of a vacuolar processing enzyme in seed.

3. The soybean plant of 1, wherein said soybean plant is genetically modified to reduce or eliminate the proteolytic activity of two or more vacuolar processing enzymes in its seed.

4. The plant of 3, wherein the plant is genetically modified to reduce or eliminate the proteolytic activity of three or more vacuolar processing enzymes in its seed.

5. The plant of 4, wherein the plant is genetically modified to inhibit the expression of four or more vacuolar processing enzymes in its seed.

6. The plant of 1, wherein at least one vacuolar processing enzyme is selected from the group consisting of Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a, and Vpe3b.

7. The plant of 8, wherein at least one vacuolar processing enzyme is selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14.

8. The plant of 1, wherein said soybean plant is stably transformed with at least one expression cassette comprising a polynucleotide encoding a polypeptide that inhibits the proteolytic activity of a vacuolar processing enzyme in seed.

9. The plant of 8, wherein said polypeptide that inhibits the proteolytic activity of a vacuolar processing enzyme is an antibody that binds to one or more soybean vacuolar processing enzymes.

10. The plant of 8, wherein said polypeptide that inhibits the proteolytic activity of a vacuolar processing enzyme is a polypeptide that specifically inhibits the activity of one or more vacuolar processing enzymes.

11. The plant of 1, wherein at least one of said seed storage proteins is selected from the group consisting of globulins and albumins.

12. The plant of 11, wherein at least one of said seed storage proteins is glycinin.

13. Transgenic seed of the plant of 1.

14. A method for producing a soybean seed storage protein having one or more altered functional properties, said method comprising the steps of

(a) transforming a soybean plant cell with at least one expression cassette capable of expressing a polynucleotide that reduces or eliminates the activity of at least one vacuolar processing enzyme in the seed of said soybean plant;

(b) regenerating a transformed plant from the transformed plant cell of step a); and

(c) collecting seed from the transformed plant of step (b).

15. The method of 14, wherein the activity of at least two vacuolar processing enzymes is reduced or eliminated in the seed of said plant.

16. The method of 15, wherein the activity of at least two vacuolar processing enzymes is reduced or eliminated in the seed of said plant.

17. The method of 16, wherein the activity of at least two vacuolar processing enzymes is reduced or eliminated in the seed of said plant.

18. The method of 14, wherein at least one vacuolar processing enzyme is selected from the group consisting of Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a, and Vpe3b.

19. The method of 18, wherein at least one vacuolar processing enzyme is selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14.

20. The method of 14, wherein at least one altered functional property is solubility of the seed storage protein.

21. The method of 20, wherein the solubility of at least one seed storage protein is increased at low pH.

22. The method of 21, wherein the solubility of the seed storage protein is increased between pH 4.0 and 6.0.

23. The method of 14, wherein at least one seed storage protein is selected from the group consisting of glycinin and 2S-albumin.

24. The method of 23, wherein said seed storage protein is glycinin.

25. The method of 14, wherein the expression cassette capable of expressing a polynucleotide that reduces or eliminates the activity of a vacuolar processing enzyme in the seed of said soybean plant comprises:

(a) a sense sequence consisting of at least 19 nucleotides corresponding to an mRNA encoding a soybean vacuolar processing enzyme; and

(b) a complementary nucleotide sequence having at least 94% identity to the complement of the sense sequence of (a).

26. The method of 25, wherein the expression cassette capable of expressing a polynucleotide that reduces or eliminates the activity of a vacuolar processing enzyme in the seed of said soybean plant comprises a loop sequence operably linked to the sense sequence and the complementary nucleotide sequence.

27. The method of 26, wherein said loop sequence additionally comprises an intron that is capable of being spliced in a soybean seed.

28. The method of 25, wherein said soybean vacuolar processing enzyme is selected from the group consisting of Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a, and Vpe3b.

29. The method of 28, wherein said sense sequence consists of at least 19 nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, and SEQ ID NO:13.

30. The method of 14, wherein the expression cassette capable of expressing a polynucleotide that reduces or eliminates the activity of a vacuolar processing enzyme in the seed of said soybean plant comprises a sense sequence consisting of at least 19 nucleotides corresponding to a messenger RNA encoding a soybean vacuolar processing enzyme.

31. The method of 30, wherein said soybean vacuolar processing enzyme is selected from the group consisting of Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a, and Vpe3b.

32. The method of 31, wherein said sense sequence consists of at least 19 nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, and SEQ ID NO:13.

33. The method of 30, wherein said soybean plant is stably transformed to express an complementary nucleotide sequence having at least 94% identity to the complement of the sense sequence.

34. The method of 33, wherein said sense sequence and said complementary nucleotide sequence are comprised within the same expression cassette.

35. The method of 33, wherein said sense sequence and said complementary nucleotide sequence are comprised within different expression cassettes.

36. The method of 14, wherein the expression cassette capable of expressing a polynucleotide that reduces or eliminates the activity of a vacuolar processing enzyme in the seed of said soybean plant comprises a complementary nucleotide sequence having at least 94% identity to the complement of a sense sequence consisting of at least 19 nucleotides of a DNA sequence corresponding to a messenger RNA for a soybean vacuolar processing enzyme.

37. The method of 14, wherein the expression cassette capable of expressing a polynucleotide that reduces or eliminates the activity of a vacuolar processing enzyme in the seed of said soybean plant comprises:

(a) a sense sequence consisting of at least 50 nucleotides of a sequence that is not endogenously expressed in soybean.

(b) a complementary nucleotide sequence having at least 94% identity to the complement of the sense sequence of (a); and

(c) a loop sequence positioned on the 3' end of the sense sequence and the 5'end of the complementary nucleotide sequence, wherein the loop sequence comprises at least 50 contiguous nucleotides corresponding to a messenger RNA encoding a soybean vacuolar processing enzyme.

38. A transformed soybean plant produced according to the method of 14.

39. A composition comprising at least one soybean seed storage protein produced according to the method of 14.

40. A method for producing a soybean seed storage protein having one or more altered functional properties, said method comprising the steps of

(a) transforming a soybean plant cell with at least one expression cassette comprising a polynucleotide encoding a polypeptide that reduces or eliminates the activity of at least one vacuolar processing enzyme in seed.

(b) regenerating a transformed plant from the transformed plant cell of step a); and

(c) collecting seed from the transformed plant of step (b).

41. The method of 40, wherein said polypeptide that inhibits the enzymatic activity of a vacuolar processing enzyme is an antibody that binds to one or more soybean vacuolar processing enzymes.

42. The method of 40, wherein said polypeptide that inhibits the enzymatic activity of a vacuolar processing enzyme is a polypeptide that inhibits the proteolytic activity of one or more soybean vacuolar processing enzymes.

42. A transformed soybean plant produced according to the method of 39.

43. A composition comprising at least one soybean seed storage protein produced according to the method of 39.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the solubility properties of legumin-type globulin protein isolated from mature wild-type and *vpe*-quad *Arabidopsis* seeds. Legumin-type globulin was isolated from sucrose density gradients. Solubility of protein obtained from these fractions was determined under low ionic strength conditions at various pH. Following incubation of the protein sample at a given pH, the amount of protein remaining in solution was quantified and graphed as a percent of the total protein added to the reaction. The error bars show standard deviations (3 replications) at each data point.

Figure 2 shows the solubility profiles for normally processed glycinin (Native Gly 11S) isolated from soybean seed and of the unprocessed proglycinin protein, obtained by expression of an appropriate expression construct in bacterial cells. The unprocessed glycinin pro-protein has much greater solubility than the native (processed) glycinin between pH 4.5 and pH 5.5.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides methods and compositions useful for altering the functional properties of soybean seed storage proteins. It is the novel finding of the present invention that the functional properties of seed storage proteins can be altered by reducing the expression and/or activity of one or more vacuolar processing enzymes in plant seed. Accordingly, the invention provides methods for altering the properties of soybean seed storage proteins by reducing or eliminating the activity of one or more endogenous vacuolar processing enzymes in soybean seed, soybean plants with altered functional properties for one or more seed storage proteins, and compositions comprising soybean seed storage proteins produced by the methods of the invention.

In some embodiments, the method comprises the steps of transforming a soybean plant cell with at least one expression cassette capable of expressing a polynucleotide that reduces of eliminates the activity of at least one soybean vacuolar processing enzyme, regenerating a transformed plant from the transformed plant cell, and collecting seed from the regenerated transformed plant.

In additional embodiments, the method comprises the steps of transforming a soybean plant cell with at least one expression cassette comprising a polynucleotide encoding a polypeptide that reduces of eliminates the activity of at least one soybean vacuolar processing enzyme, regenerating a transformed plant from the transformed plant cell, and collecting seed from the regenerated transformed plant. The seed harvested from the transformed plant contains seed storage proteins having altered functional properties.

The invention also provides soybean seed storage proteins having altered functional properties, and compositions comprising these storage proteins.

Also provided are plants that are genetically modified or mutagenized to reduce or eliminate the activity of one or more soybean vacuolar processing enzymes, and transformed seed of these plants.

The methods and compositions of the invention are described in more detail below.

SOYBEAN SEED STORAGE PROTEINS

The invention relates to methods of altering the functional properties of one or more seed storage proteins in soybean, and to soybean plants that are genetically modified or mutagenized to alter the functional properties of one or more seed storage proteins. The functional properties of any soybean seed storage protein may be altered according to the invention. Soybean has three major seed storage proteins; two globulins, glycinin (also known as the 11S globulins) and β-conglycinin (also known as the 7S globulins), and one albumin, 2S albumin. Together, these proteins comprise 70% to 80% of the soybean seed's total protein, or 25 to 35% of the seed's dry weight. Glycinin is a large protein with a molecular weight of about 360 kDa. It is a hexamer composed of the various combinations of five different types of subunits, which are identified as G1, G2, G3, G4 and G5. Each subunit is composed of one acidic region and one basic region held together by a disulfide bond. The glycinin subunits are primarily encoded by genes designated *Gy1*, *Gy2*, *Gy3*, *Gy4* and *Gy5*, corresponding to subunits G1, G2, G3, G4 and G5, respectively (Nielsen, N. C. *et al.* (1989) *Plant Cell* 1:313-328). At least one other gene, *Gy7*, also appears to encode a glycinin subunit (Beilinson *et al.* (2002) *Theor. Appl. Genet.* 104:1132-40).

β.-conglycinin is a heterogeneously glycosylated protein with a molecular weight ranging from 150 and 240 kDa. It is composed of varying combinations of three highly negatively charged subunits identified as α, α', and β. The three classes of β-conglycinin subunits are encoded by a total of 15 subunit genes clustered in several regions within the genome soybean (Harada, J. J. *et al.* (1989) *Plant Cell* 1:415-425).

The sulfur-rich 2S albumin comprises between 5-10% of the soybean seed's total protein. See, NCBI Accession No. AF005030, U.S. Patent No. 5,850,016, and Alfredo *et al.* (1997) *Plant Physiol.* 114: 1567, each of which is herein incorporated by reference.

Over the past 20 years, significant effort has been aimed at understanding the functional properties of soybean seed storage proteins. See, for example, Kinsella *et al.*

(1985) *New Protein Foods* 5:107-179; Morr (1987) *JAOCS* 67:265-271; and Peng *et al.* (1984) *Cereal Chem.* 61:480-489. Examples of functional properties of interest include solubility, water adsorption, binding, and retention, gelation (including gel firmness), cohesion-adhesion, elasticity, emulsification, fat-adsorption, flavor binding, foaming, and color control. See, for example, Kinsella (1979) *J. Amer. Oil Chemists Soc.* 56:242-58, herein incorporated by reference. The present invention relates the alteration of the functional properties of soybean seed storage proteins, such as the solubility, water retention properties, gelation properties, or emulsification properties of soybean seed storage proteins. These functional properties are related, and thus an alteration in one functional property (such as solubility) can lead to an alteration in other functional properties. Thus, in some embodiments, one functional property is altered, while in other embodiments, multiple functional properties such as two or more functional properties, three or more functional properties, or four or more functional properties are altered.

In some embodiments, the gelation properties of one or more soybean storage proteins are altered. By "gelation properties" it is intended the ability of a protein to form a three-dimensional matrix of intertwined, partially associated polypeptides in which water can be held. See, for example, Kinsella (1979) *J. Amer. Oil Chemists Soc.* 56:242-58, herein incorporated by reference.

In some embodiments, the emulsification properties of one or more soybean storage proteins are altered. By "emulsification properties" it is intended the ability of a protein to aid in the uniform formation and stabilization of fat emulsions. See, for example, Kinsella (1979) *J. Amer. Oil Chemists Soc.* 56:242-58, herein incorporated by reference.

In some embodiments, the water retention properties of one or more soybean storage proteins is altered. Water retention of soybean protein isolates is dependent in part on the proteolyzed state of the proteins in the isolate (Mietsch *et al.* (1989) *Nahrung* 33:9-15).

In some embodiments, the solubility of one or more soybean seed storage proteins is altered. By "solubility" it is intended dispersibility in fluid. Solubility may be measured using the nitrogen solubility index (NSI) or the protein dispersibility index.

See, Johnson (1970) *Food. Prod. Dev.* 3:78, and Johnson (1970) *JAOCS* 47:402; both of which are herein incorporated by reference in their entireties. The solubility of a protein solution can be measured by centrifuging the solution at 17,000 × g for 10 minutes, and then assaying the supernatant to determine protein content.

It is the novel finding of the present invention that eliminating the expression of vacuolar processing enzymes in seed results in a marked alteration in the solubility of seed storage proteins. The legumin-like seed storage proteins of *Arabidopsis* are relatively insoluble at low pH, having less than 20% solubility in solutions having a pH of less than 5, and only about 25% solubility at pH 5.5. However, in an *Arabidopsis* plant null for α, β, γ, and δ vacuolar processing enzymes, the legumin-type globulin proteins show greatly enhanced solubility between pH 3.5 and pH 5.0. See Figure 1 and the Experimental section.

The present invention also shows that soybean glycinin proteins that are not cleaved by vacuolar processing enzymes have increased solubility at low pH in comparison with glycinin that is cleaved by vacuolar processing enzymes. *See*, Figure 2. Accordingly, reducing the expression of soybean vacuolar processing enzymes increases the solubility of glycinin in soybean seed.

Thus, in some embodiments, the present invention provides methods of producing a soybean seed storage protein having increased solubility, and soybean plants that have been genetically modified to increase the solubility of a seed storage protein. A seed storage protein in a plant that has been genetically modified to inhibit the expression of one or more vacuolar processing enzymes has increased solubility according to the invention if the solubility of the protein is at least 2 times greater than the solubility of the same protein in a plant that has not been genetically modified to inhibit the expression of a vacuolar processing enzyme. In some embodiments, the solubility of the soybean seed storage protein in a plant that has been genetically modified to inhibit the expression of one or more vacuolar processing enzymes is at least 5 times greater than, at least 10 times greater than, at least 20 times greater than, at least 50 times greater than, at least 100 times greater than, or more than 100 times greater than the solubility of the same protein

in a plant that has not been genetically modified to inhibit the expression of a vacuolar processing enzyme.

In some embodiments of the invention, the solubility of a seed storage protein is increased at low pH. For example, the invention provides embodiments in which the solubility of the seed storage protein is increased in the pH range between pH 3.5 and pH 6.5. In particular embodiments, the solubility of the seed storage protein is increased between pH 4.0 and 6.0, such as between pH 4.5 and 5.5. Soybean seed storage proteins having increased solubility according to the invention will be at least 10% soluble, at least 20% soluble, at least 30% soluble, at least 40% soluble, at least 50% soluble, at least 60% soluble, at least 70% soluble, at least 80% soluble, or more than 80% soluble in solutions having a pH ranging between 4.5 and 5.5. In some embodiments, one or more of the seed storage proteins is glycinin. In another embodiment one or more of the seed storage proteins is 2S albumin.

The invention also encompasses soybean seed storage proteins having altered functional properties, and compositions comprising these seed storage proteins. Soy protein products are generally categorized into three major groups: soy flours and grits containing about 45 to 54% soy protein on a moisture free basis; soy protein concentrates containing 65 to 90% protein on a moisture free basis; and soy protein isolates having a minimum of 90% protein on a moisture free basis. Soy protein isolates are preferred in many applications because of their higher protein content, easier digestibility, and improved flavor as compared with soy flours, grits and concentrates. In one embodiment, the invention pertains to the production of soy protein isolates, which are the most highly refined soy protein products commercially available.

SOYBEAN VACUOLAR PROCESSING ENZYMES

According to the invention, the proteolytic activity of at least one, at least two, at least three, at least four, at least five, at least six, or at least seven, or more than seven vacuolar processing enzymes may be reduced or eliminated in soybean seed. In plants, vacuolar processing enzymes (VPE's) comprise a small gene family of plant asparaginyl endopeptidases implicated in the control of several important cellular processes including

storage protein proteolysis involved in protein turnover and mobilization of amino acid reserves in vegetative tissue during plant senescence process. See, for example, Hara-Nishimura *et al.* (1987) *Plant Physiol* 85:440-445; D'Hondt *et al.* (1993) *J. Biol. Chem.* 268:20884-20891; Hara-Nishimura *et al.* (1993) *Plant Cell* 5:1651-1659; Hara-Nishimura *et al.* (1995) *J. Plant Physiol.* 145:632-640; and Kinoshita *et al.* (1995) *Plant Cell Physiol.* 36:1555-1562; D'Hondt *et al.* (1997) *Plant Molec. Biol.* 33:187-192; Barrett *et al.*, ed. (1998) *Handbook of Proteolytic Enzymes*, Academic Press, Sand Diego, pp746-749, each of which is incorporated by reference.

Vacuolar processing enzymes are a member of peptidase family C13 (see Pfam Accession number PF01650), and catalyze the hydrolysis of proteins at -Asn-|-Xaa peptide bonds. These cysteine proteases are members of enzyme class 3.4.22.34. Alternate names for this family include legumain, asparaginyl endopeptidase, phaseolin endopeptidase, and bean endopeptidase. This family of peptidases is described, for example, in Hara-Nishimura, Asparinyl endopeptidase in *Handbook of Proteolytic Enzymes*, Barrett *et al.*, eds., pp. 746 -749 (1998) Academic Press, London; Dalton and Brindley, Schistosome Legumain in *Handbook of Proteolytic Enzymes*, Barrett *et al.*, eds., pp. 749-754 (1998) Academic Press, London; Chen *et al.* (1998) *FEBS Letters* 441:361-65, and Muntz and Shutov (2002) *Trends in Plant Science* 7:340-44; each of which is hérein incorporated by reference.

By a "soybean vacuolar processing enzyme" as used herein, it is intended a soybean cysteine protease that is a member of the peptidase C13 family (Pfam Accession number PF01650) and has the proteolytic activity of enzyme class 3.4.22.34, i.e. the ability catalyze the hydrolysis of proteins at -Asn-|-Xaa- peptide bonds. See Chen *et al.* (1998) *FEBS Letters* 441:361-365 for a description of active site residues involved in vacuolar processing enzyme activity. See Jung *et al.* (1998) *The Plant Cell* 10:343-57, herein incorporated by reference, for a description of the substrate specificity of soybean vacuolar processing enzymes in soybean and for assays for determining vacuolar processing enzyme activity.

The present invention provides amino acid sequences for soybean Vpe1a (SEQ ID NO:2), Vpe1b (SEQ ID NO:4), Vpe2a (SEQ ID NO:6), Vpe2b(SEQ ID NO:8), and

Vpe3a(SEQ ID NO:10). Nucleotide sequences encoding these soybean VPEs are set forth in SEQ ID NO:1 (Vpe1a), SEQ ID NO:3 (Vpe1b), SEQ ID NO:5 (Vpe2a), SEQ ID NO:7 (Vpe2b), and SEQ ID NO:9 (Vpe3a).

Soybean vacuolar processing enzymes (VPE's) have been also described in the art. See, for example, the soybean VPE described by Shimada *et al.* (1994) *Plant Cell Physiol.* 35:713-718. The coding sequence for this soybean VPE is set forth as SEQ ID NO:11, and the encoded protein is set forth in SEQ ID NO:12. See also NCBI Accession number AF169019. The coding sequence for this soybean VPE is set forth as SEQ ID NO:13, and the encoded protein is set forth in SEQ ID NO:14.

The soybean VPE's can be grouped phylogentically into gene sub families, as has been described for members of the VPE gene family of other plants (Muntz and Shutov (2002) *Trends in Plant Science* 7:340-44). Soybean Vpe1a and Vpe1b are seed-type VPE's and are closely related to β-VPE from *Arabidopsis*, while Vpe2a, Vpe2b, Vpe3a, and Vpe3b are vegetative-type VPE's and closely related to α- and γ-VPE from Arabidopsis.

Thus, in some embodiments of the invention, at least one of the vacuolar processing enzymes whose activity is reduced is selected from the group consisting of Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a, and Vpe3b. In further embodiments, at least one vacuolar processing enzyme whose expression is inhibited is selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, and SEQ ID NO:14.

The invention encompasses the inhibition of the expression of soybean homologs of the proteins set forth in SEQ ID NOS:2, 4, 6, 8, 10, 12, and 14. Such soybean homologs typically have substantial sequence similarity with at least one amino acid sequence selected from SEQ ID NOS:2, 4, 6, 8, 10, 12, and 14, and the nucleotide sequences encoding them typically have substantially similarity to at least one nucleotide sequence selected from SEQ ID NOS; 1, 3, 5, 7, 9, 11, and 13. The homologs also have the protease activity of a protein set forth in SEQ ID NO:2, 4, 6, 8, 10, 12, or 14, *i.e.*, the homologs catalyze the hydrolysis of proteins at -Asn-|-Xaa- peptide bonds. Thus in some embodiments, the invention comprises inhibiting the expression of a soybean vacuolar

protease encoded by a sequence having at least 70% sequence identity, at least 80% sequence identity, at least 85% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity, at least 99% sequence identity, or more than 99% sequence identity with at least one nucleotide sequence selected from SEQ ID NOS; 1, 3, 5, 7, 9, 11, and 13. Methods of calculating the level of sequence identity between two sequences are provided elsewhere herein.

The proteolytic activity of a soybean vacuolar processing enzyme may determined by any method known in the art. Methods for determining the proteolytic activity of a vacuolar processing enzyme are described, for example, in Jung *et al.* (1998) *The Plant Cell* 10:343-57, Hara-Nishimura, Asparinyl endopeptidase in *Handbook of Proteolytic Enzymes*, Barrett *et al.*, eds., pp. 746 -749 (1998) Academic Press, London; and Dalton and Brindley, Schistosome Legumain in *Handbook of Proteolytic Enzymes*, Barrett *et al.*, eds., pp. 749-754 (1998) Academic Press, London; Chen *et al.* (1998) *FEBS Letters* 441:361-65;; each of which is herein incorporated by reference.

METHODS OF REDUCING THE PROTEOLYTIC ACTIVITY OF VACUOLAR PROCESSING ENZYMES

The present invention encompasses methods of producing one or more seed storage proteins having altered functional properties by reducing or eliminating the proteolytic activity of one or more vacuolar processing enzymes. The invention also encompasses soybean plants that have been genetically modified or mutagenized to reduce or eliminate the activity of one or more vacuolar processing enzymes.

In some embodiments, the activity of the vacuolar processing enzyme is reduced or eliminated by transforming a soybean plant cell with an expression cassette that expresses a polynucleotide that inhibits the expression of the vacuolar processing enzyme. The polynucleotide may inhibit the expression of one or more vacuolar processing enzymes directly, by preventing translation of the vacuolar processing enzyme messenger RNA, or indirectly, by encoding a polypeptide that inhibits the transcription or translation of a soybean gene encoding a vacuolar processing enzyme. Methods for

inhibiting or eliminating the expression of a gene in a plant are well known in the art, and any such method may be used in the present invention to inhibit the expression of one or more soybean vacuolar processing enzymes.

The expression of a vacuolar processing enzyme is inhibited according to the present invention if the protein level of the vacuolar processing enzyme is less than 70% of the protein level of the same vacuolar processing enzyme in a plant that that has not been genetically modified or mutagenized to inhibit the expression of that vacuolar processing enzyme. In particular embodiments of the invention, the protein level of the vacuolar processing enzyme in a modified plant according to the invention is less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, or less than 5% than of the protein level of the same vacuolar processing enzyme in a plant that this is not a mutant or that has not been genetically modified to inhibit the expression of that vacuolar processing enzyme. The expression level of the vacuolar processing enzyme may be measured directly, by assaying for the level of vacuolar processing enzyme expressed in the soybean cell or plant, or indirectly, by measuring the proteolytic activity of the vacuolar processing enzyme in the soybean cell or plant. Methods for determining the proteolytic activity of vacuolar processing enzymes are described elsewhere herein.

In other embodiments of the invention, the activity of one or more soybean vacuolar processing enzymes is reduced or eliminated by transforming a soybean plant cell with an expression cassette comprising a polynucleotide encoding a polypeptide that inhibits the activity of one or more soybean vacuolar processing enzymes. The proteolytic activity of a vacuolar processing enzyme is inhibited according to the present invention if the proteolytic activity of the vacuolar processing enzyme is less than 70% of the proteolytic activity of the same vacuolar processing enzyme in a plant that has not been genetically modified to inhibit the proteolytic activity of that vacuolar processing enzyme. In particular embodiments of the invention, the proteolytic activity of the vacuolar processing enzyme in a modified plant according to the invention is less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, or less than 5% than of the proteolytic activity of the same vacuolar processing enzyme in a

plant that that has not been genetically modified to inhibit the expression of that vacuolar processing enzyme. The proteolytic activity of a vacuolar processing enzyme is "eliminated" according to the invention when it is not detectable by the assay methods described elsewhere herein. Methods of determining the proteolytic activity of a vacuolar processing enzyme are described elsewhere herein.

In other embodiments, the activity of a vacuolar processing enzyme may be reduced or eliminated by disrupting the gene encoding the vacuolar processing enzyme. The invention encompasses mutagenized soybean plants that carry mutations in VPE genes, where the mutations reduce expression of the VPE genes or inhibit the proteolytic activity of the encoded VPE.

Thus, many methods may be used to reduce or eliminate the activity of a vacuolar processing enzyme. More than one method may be used to reduce the activity of a single soybean vacuolar processing enzyme. In addition, combinations of methods may be employed to reduce or eliminate the activity of two or more different vacuolar processing enzymes, three or more different vacuolar processing enzymes, four or more different vacuolar processing enzymes, five or more different vacuolar processing enzymes, or six or more different vacuolar processing enzymes.

Non-limiting examples of methods of reducing or eliminating the expression of a soybean vacuolar processing enzyme are given below.

I. Polynucleotides That Inhibit the Expression of One or More Vacuolar Processing Enzymes

In some embodiments of the present invention, a soybean plant cell is transformed with an expression cassette that is capable of expressing a polynucleotide that inhibits the expression of one or more vacuolar processing enzymes. The term "expression" as used herein refers to the biosynthesis of a gene product, including the transcription and/or translation of said gene product. For example, for the purposes of the present invention, an expression cassette capable of expressing a polynucleotide that inhibits the expression of at least one soybean vacuolar processing enzyme is an expression cassette capable of producing an RNA molecule that inhibits the transcription and/or translation of at least

one soybean vacuolar processing enzyme. The "expression" or "production" of a protein or polypeptide from a DNA molecule refers to the transcription and translation of the coding sequence to produce the protein or polypeptide, while the "expression" or "production" of a protein or polypeptide from an RNA molecule refers to the translation of the RNA coding sequence to produce the protein or polypeptide.

Examples of polynucleotides that inhibit the expression of a soybean vacuolar processing enzyme are given below.

A. Sense Suppression/Cosuppression

In some embodiments of the invention, inhibition of the expression of a vacuolar processing enzyme may be obtained by sense suppression or cosuppression. For cosuppression, the expression cassette is designed to express an RNA molecule corresponding to all or part of a messenger RNA encoding a soybean vacuolar processing enzyme in the "sense" orientation. Over expression of the RNA molecule can result in reduced expression of the native gene. Accordingly, multiple plants lines transformed with the cosuppression expression cassette are screened to identify those that show the greatest inhibition of vacuolar processing enzyme expression.

The polynucleotide used for cosuppression may correspond to all or part of the sequence encoding the vacuolar processing enzyme, all or part of the 5' and/or 3' untranslated region of a vacuolar processing enzyme transcript, or all or part of both the coding sequence and the untranslated regions of a transcript encoding a vacuolar processing enzyme. In some embodiments where the polynucleotide comprises all or part of the coding region of the vacuolar processing enzyme, the expression cassette is designed to eliminate the start codon of the polynucleotide so that no protein product will be transcribed.

Cosuppression may be used to inhibit the expression of plant genes to produce plants having undetectable protein levels for the proteins encoded by these genes. See, for example, Broin *et al.* (2002) *The Plant Cell* 14:1417-32. Cosuppression may also be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Methods for using cosuppression to inhibit the expression of

endogenous genes in plants are described in Flavell *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3490-96; Jorgensen *et al.* (1996) *Plant Molec. Biol.* 31:957-73; Johansen and Carrington (2001) *Plant Physiology* 126:930-938; Broin *et al.* (2002) *The Plant Cell* 14:1417-1432; Stoutjesdijk *et al* (2002) *Plant Physiology* 129:1723-1731; Yu *et al.* (2003) *Phytochemistry* 63:753-63; and U.S. Patent Nos. 5,034,323, 5,283,184, and 5,942,657; each of which is herein incorporated by reference. The efficiency of cosuppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the sense sequence and 5' of the polyadenylation signal. *See*, U.S. Patent Publication 20020048814, herein incorporated by reference.

B. Antisense Suppression

In some embodiments of the invention, inhibition of the expression of a vacuolar processing enzyme may be obtained by antisense suppression. For antisense suppression, the expression cassette is designed to express an RNA molecule complementary to all or part of a messenger RNA encoding a soybean vacuolar processing enzyme. Overexpression of the antisense RNA molecule can result in reduced expression of the native gene. Accordingly, multiple plants lines transformed with the antisense suppression expression cassette are screened to identify those that show the greatest inhibition of vacuolar processing enzyme expression.

The polynucleotide for use in antisense suppression may correspond to all or part of the complement of the sequence encoding the vacuolar processing enzyme, all or part of the complement of the 5' and/or 3' untranslated region of a vacuolar processing enzyme transcript, or all or part of the complement of both the coding sequence and the untranslated regions of a transcript encoding a vacuolar processing enzyme. In addition, the antisense polynucleotide may be fully complementary (i.e. 100% identical to the complement of the target sequence) or partially complementary (i.e. less than 100% identical to the complement of the target sequence) to the target sequence. Antisense suppression may be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Methods for using antisense suppression to inhibit the expression of endogenous genes in plants are described, for example, in Liu *et*

al (2002) *Plant Physiology* 129:1732-43 and U.S. Patent Nos. 5,759,829 and 5,942,657, each of which is herein incorporated by reference. Efficiency of antisense suppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the antisense sequence and 5' of the polyadenylation signal. *See*, U.S. Patent Publication 20020048814, herein incorporated by reference.

C. Double Stranded RNA Interference

In some embodiments of the invention, inhibition of the expression of a vacuolar processing enzyme may be obtained by double stranded RNA (dsRNA) interference. For dsRNA interference, a sense RNA molecule like that described above for cosuppression and an antisense RNA molecule that is fully or partially complementary to the sense RNA molecule are expressed in the same cell, resulting in inhibition of the expression of the corresponding endogenous messenger RNA.

Expression of the sense and antisense molecules can be accomplished by designing the expression cassette to comprise both a sense sequence and an antisense sequence. Alternatively, separate expression cassettes may be used for the sense and antisense sequences. Multiple plants lines transformed with the dsRNA interference expression cassette or expression cassettes are then screened to identify plant lines that show the greatest inhibition of vacuolar processing enzyme expression. Methods for using dsRNA interference inhibit the expression of endogenous plant genes are described in Waterhouse *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-64, Liu *et al.* (2002) *Plant Physiology* 129:1732-43, and WO publications WO9949029, WO9953050, WO9961631, and WO049035; each of which is herein incorporated by reference.

D. Hairpin RNA Interference and Intron-Containing Hairpin RNA Interference

In some embodiments of the invention, inhibition of the expression of one or more vacuolar processing enzyme may be obtained by hairpin RNA (hpRNA) interference or intron-containing hairpin RNA (ihpRNA) interference. These methods

are highly efficient at inhibiting the expression of endogenous genes. *See*, Waterhouse and Helliwell (2003) *Nat. Rev. Gen.* 4:29-38 and the references cited therein.

For hpRNA interference, the expression cassette is designed to express an RNA molecule that hybridizes with itself to form a hairpin structure that comprises a single-stranded loop region and a base-paired stem. The base-paired stem region comprises a sense sequence corresponding to all or part of the endogenous messenger RNA encoding the gene whose expression is to be inhibited, and an antisense sequence that is fully or partially complementary to the sense sequence. Thus, the base-paired stem region of the molecule generally determines the specificity of the RNA interference. hpRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and the RNA interference they induce is inherited by subsequent generations of plants. See, for example, Chuang and Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-90; Stoutjesdijk *et al.* (2002) *Plant Physiology* 129:1723-31; and Waterhouse and Helliwell (2003) *Nat. Rev. Gen.* 4:29-38. Methods for using hpRNA interference to inhibit or silence the expression of genes are described, for example, in Chuang and Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-90; Stoutjesdijk *et al.* (2002) *Plant Physiology* 129:1723-31; Waterhouse and Helliwell (2003) *Nat. Rev. Gen.* 4:29-38; Pandolfini *et al. BMC Biotechnology* 3:7, and U.S. Patent Publication 20030175965, each of which is herein incorporated by reference. A transient assay for the efficiency of hpRNA constructs to silence gene expression *in vivo* has been described by Panstruga *et al.* (2003) *Mol. Biol. Rep.* 30:135-40, herein incorporated by reference.

For ihpRNA, the interfering molecules have the same general structure as for hpRNA, but the RNA molecule additionally comprises an intron that is capable of being spliced in the cell in which the ihpRNA is expressed. The use of an intron minimizes the size of the loop in the hairpin RNA molecule following splicing, and this increase the efficiency of interference. See, for example, Smith *et al.* (2000) *Nature* 407:319-320. In fact, Smith *et al.* show 100% suppression of endogenous gene expression using ihpRNA-mediated interference. Methods for using ihpRNA interference to inhibit the expression of endogenous plant genes are described, for example, in Smith *et al.* (2000) *Nature* 407:319-320; Wesley *et al.* (2001) *The Plant Journal* 27:581-590; Wang and Waterhouse

(2001) *Current Opinion in Plant Biology* 5:146-150; Waterhouse and Helliwell (2003) *Nat. Rev. Gen.* 4:29-38; Helliwell and Waterhouse (2003) *Methods.* 30:289-95, and U.S. Patent Publication No. 20030180945, each of which is herein incorporated by reference.

The expression cassette for hpRNA interference may also be designed such that the sense sequence and the antisense sequence do not correspond to an endogenous RNA. In this embodiment, the sense and antisense sequence flank a loop sequence that comprises a nucleotide sequence corresponding to all or part of the endogenous messenger RNA of the target gene. Thus, it is the loop region that determines the specificity of the RNA interference. See, for example, patent publication WO 0200904, herein incorporated by reference.

E. Amplicon-Mediated Interference

Amplicon expression cassettes comprise a plant virus-derived sequence that contains all or part of the target gene, but generally not all of the genes of the native virus. The viral sequences present in the transcription product of the expression cassette allow the transcription product direct its own replication. The transcripts produced by the amplicon may be either sense or antisense relative to the target sequence (i.e. the messenger RNA for a soybean vacuolar processing enzyme). Methods of using amplicons to inhibit the expression of endogenous plant genes are described, for example, in Angell and Baulcombe (1997) *EMBO J.* 16:3675-84, Angell and Baulcombe (1999) *The Plant Journal* 20:357-362, and U.S. Patent No. 6,646,805, each of which is herein incorporated by reference.

F. Ribozymes

In some embodiments, the polynucleotide expressed by the expression cassette of the invention is catalytic RNA or ribozyme activity specific for the messenger RNA of a vacuolar processing enzyme. Thus, the polynucleotide causes the degradation of the endogenous messenger RNA, resulting in reduced expression of the vacuolar processing enzyme. This method is described, for example, in U.S. Patent No. 4,987,071, herein incorporated by reference.

G. Small interfering RNA or micro RNA

In some embodiments of the invention, inhibition of the expression of one or more vacuolar processing enzyme may be obtained by RNA interference by expression of a gene encoding a micro RNA (miRNA). miRNAs are regulatory agents consisting of about 22 ribonucleotides. miRNA are highly efficient at inhibiting the expression of endogenous genes. *See*, for example Javier *et al.* (2003) *Nature* 425: 257–263; herein incorporated by reference.

For miRNA interference, the expression cassette is designed to express an RNA molecule that is modeled on an endogenous miRNA gene. The miRNA gene encodes an RNA that forms a hairpin structure containing a 22nt sequence that is complementary to another endogenous gene (target sequence). For suppression of VPE expression the 22nt sequence is selected from a VPE transcript sequence and contains 22 nt of said soybean VPE sequence in sense orientation and 21nt of an corresponding antisense sequence that is complementary to the sense sequence. miRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and the RNA interference they induce is inherited by subsequent generations of plants.

II. Polypeptides that Inhibit the Expression of Vacuolar Processing Enzymes

In some embodiments, the present invention provides a method for producing a soybean seed storage protein having one or more altered functional properties, where the method comprises the steps of transforming a soybean plant cell with at least one expression cassette comprising a polynucleotide encoding a polypeptide that inhibits the expression of one or more soybean vacuolar processing enzymes, regenerating a transformed plant from the transformed plant cell, and collecting seed from the transformed plant. The polynucleotide may encode any polypeptide that inhibits the expression of a soybean vacuolar processing enzyme.

In one embodiment, the polynucleotide encodes a zinc finger protein that binds to a gene encoding a soybean vacuolar processing enzyme, resulting in reduced expression of the gene. In particular embodiments, the zinc finger protein binds to a regulatory

region of a vacuolar processing enzyme gene. In other embodiments, the zinc finger protein binds to a messenger RNA encoding a vacuolar processing enzyme and prevents its translation. Methods of selecting sites for targeting by zinc finger proteins have been described, for example, by U.S. Patent No. 6,453,242, herein incorporated by reference. Methods for using zinc finger proteins to inhibit the expression of genes in plants are described, for example, in U.S. Patent Publication 20030037355, herein incorporated by reference.

III. Polypeptides that Inhibit the Proteolytic Activity of Vacuolar Processing Enzymes

In some embodiments, the present invention provides a method for producing a soybean seed storage protein having one or more altered functional properties, where the method comprises the steps of transforming a soybean plant cell with at least one expression cassette comprising a polynucleotide encoding a polypeptide that inhibits the proteolytic activity of one or more soybean vacuolar processing enzymes, regenerating a transformed plant from the transformed plant cell, and collecting seed from the transformed plant. The polynucleotide may encode any polypeptide that inhibits the activity of a soybean vacuolar processing enzyme.

In some embodiments of the invention, the polynucleotide encodes an antibody that binds to at least one soybean VPE, and reduces the proteolytic activity of the VPE. In another embodiment, the binding of the antibody results in increased turn-over of the antibody-VPE complex by cellular quality control mechanisms. The expression of antibodies in plant cells and the inhibition of molecular pathways by expression and binding of antibodies to proteins in plant cells are well known in the art. See, for example, Conrad and Sonnewald (2003) *Nature Biotech.* 21:35–36, incorporated herein by reference.

In other embodiments of the invention, the polynucleotide encodes a polypeptide that specifically inhibits the proteolytic activity of a soybean vacuolar processing enzyme, i.e. a proteinase inhibitor. In particular embodiments, the proteinase inhibitor is a C-terminal propeptide of a VPE that functions as an auto-inhibitory domain. See, for

example, Kuroyangi *et al.* (2002) *Plant Cell Physiol.* 43:143-151, herein incorporated by reference. The expression of other proteinase inhibitors in plant cells is well known in the art. See, for example, Zhong *et al.* (1999) *Molecular Breeding* 5: 345-56, herein incorporated by reference.

IV. Methods of Disrupting a Gene Encoding a Soybean Vacuolar Processing Enzyme

In some embodiments of the present invention, the activity of a vacuolar processing enzyme is reduced or eliminated by disrupting the gene encoding the vacuolar processing enzyme. The gene encoding the vacuolar processing enzyme may be disrupted by any method know in the art. For example, in one embodiment the gene is disrupted by transposon tagging. In another embodiment, the gene is disrupted by mutagenizing soybean plants using random or targeted mutagenesis, and selecting for plants that have reduced vacuolar processing enzyme activity.

A. Transposon Tagging

In one embodiment of the invention, transposon tagging is used to reduce or eliminate the proteolytic activity of one or more soybean vacuolar processing enzymes. Transposon tagging comprises inserting a transposon within an endogenous soybean vacuolar processing enzyme gene to reduce or eliminate expression of the vacuolar processing enzyme. By "vacuolar processing enzyme gene" is meant the gene that encodes a soybean vacuolar processing enzyme according to the invention.

In this embodiment, the expression of one or more vacuolar processing enzymes is reduced or eliminated by inserting a transposon within a regulatory region or coding region of the gene encoding the vacuolar processing enzyme A transposon that is within an exon, intron, 5' or 3' untranslated sequence, a promoter, or any other regulatory sequence of a soybean vacuolar processing enzyme gene may be used to reduce or eliminate the expression and/or activity of the encoded vacuolar processing enzyme.

Methods for the transposon tagging of specific genes in plants are well known in the art. See, for example, Maes *et al.* (1999) *Trends Plant Sci.* 4:90-96; Dharmapuri and

Sonti (1999) *FEMS Microbiol. Lett.* 179:53-59; Meissner *et al.* (2000) *Plant J.* 22:265-274; Phogat et al. (2000) *J. Biosci.* 25:57-63; Walbot (2000) *Curr. Opin. Plant Biol.* 2:103-107; Gai *et al.* (2000) *Nuc. Acids Res.* 28:94-96; Fitzmaurice *et al.* (1999) *Genetics* 153:1919-1928). In addition, the TUSC process for selecting Mu insertions in selected genes has been described in Bensen *et al.* (1995) *Plant Cell* 7:75-84; Mena *et al.* (1996) *Science* 274:1537-1540; and U.S. Patent No. 5,962,764, each of which is herein incorporated by reference.

B. Mutant Soybean Plants with Reduced Activity for One or More VPEs

Additional methods for decreasing or eliminating the expression of endogenous genes in plants are also known in the art and can be similarly applied to the instant invention. These methods include other forms of mutagenesis, such as ethyl methanesulfonate-induced mutagenesis, deletion mutagenesis, and fast neutron deletion mutagenesis used in a reverse genetics sense (with PCR) to identify plant lines in which the endogenous gene has been deleted. For examples of these methods see Ohshima, *et al.* (1998) *Virology* 243:472-481; Okubara *et al.* (1994) *Genetics* 137:867-874; and Quesada *et al.* (2000) *Genetics* 154:421-436; each of which is herein incorporated by reference. In addition, a fast and automatable method for screening for chemically induced mutations, TILLING, (Targeting Induced Local Lesions In Genomes), using denaturing HPLC or selective endonuclease digestion of selected PCR products is also applicable to the instant invention. See McCallum *et al.* (2000) *Nat. Biotechnol.* 18:455-457, herein incorporated by reference.

Mutations that impact gene expression or that interfere with the function (enzymatic activity) of the encoded protein are well known in the art. Insertional mutations in gene exons usually result in null-mutants. Mutations in conserved active site residues are particularly effective in inhibiting the enzymatic activity of the encoded protein. Active site residues of plant VPE's suitable for mutagenesis with the goal to eliminate VPE enzymatic activity have been described. See, for example, Hara-Nishimura, "Asparinyl Endopeptidases" in *Handbook of Proteolytic Enzymes*, Barrett *et al.*, eds., pp. 746 -749 (1998) Academic Press, London; Dalton and Brindley,

"Schistosome Legumain" in *Handbook of Proteolytic Enzymes*, Barrett *et al.*, eds., pp. 749-754 (1998) Academic Press, London; and Chen *et al.* (1998) *FEBS Letters* 441:). Such mutants can be isolated according to well-known procedures, and mutations in different VPE loci can be stacked by genetic crossing. See, for example, Gruis *et al.* (2002) *Plant Cell* 14:2863-82.

In another embodiment of this invention, dominant mutants can be used to trigger RNA silencing due to gene inversion and recombination of a duplicated gene locus. See, for example, Kusaba *et al.* (2003) *Plant Cell* 15:1455-67.

The invention encompasses additional methods for reducing or eliminating the activity of one or more vacuolar processing enzymes. Examples of other methods for altering or mutating a genomic nucleotide sequence in a plant are known in the art and , include, but are not limited to, the use of chimeric vectors, chimeric mutational vectors, chimeric repair vectors, mixed-duplex oligonucleotides, self-complementary chimeric oligonucleotides, and recombinogenic oligonucleobases. Such vectors and methods of use, such as, for example, chimeraplasty, are known in the art. Chimeraplasty involves the use of such nucleotide constructs to introduce site-specific changes into the sequence of genomic DNA within an organism. See, for example, U.S. Patent Nos. 5,565,350; 5,731,181; 5,756,325; 5,760,012; 5,795,972; and 5,871,984; each of which are herein incorporated by reference. See also, WO 98/49350, WO 99/07865, WO 99/25821, and Beetham *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96:8774-8778; each of which is herein incorporated by reference.

EXPRESSION CASSETTES

The present invention encompasses to the transformation of soybean plants with expression cassettes capable of expressing polynucleotides that reduce or eliminate the proteolytic activity of one or more vacuolar processing enzymes. The expression cassette will include in the 5'-3' direction of transcription, a transcriptional and translational initiation region (i.e., a promoter) and a polynucleotide of interest, i.e., a polynucleotide capable of directly or indirectly (i.e. via expression of a protein product) reducing or eliminating the activity of one or more soybean vacuolar processing enzymes. The

expression cassette may optionally comprise a transcriptional and translational termination region (i.e. termination region) functional in plants. In some embodiments, the expression cassette comprises a selectable marker gene to allow for selection for stable transformants. Expression constructs of the invention may also comprise a leader sequence and/or a sequence allowing for inducible expression of the polynucleotide of interest. *See*, Guo *et al.* (2003) *Plant J.* 34:383-92 and Chen *et al.* (2003) *Plant J.* 36:731-40 for examples of sequences allowing for inducible expression.

The regulatory sequences of the expression construct will be operably linked to the polynucleotide of interest. By "operably linked" is intended a functional linkage between a promoter and a second sequence wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleotide sequences being linked are contiguous.

According to the invention, the proteolytic activity of at least one, at least two, at least three, at least four, at least five, or at least six at least seven, or more than seven vacuolar processing enzymes may be reduced or eliminated in soybean seed. In some embodiments, the polynucleotide of interest is designed to reduce or eliminate the activity of only one vacuolar processing enzyme, while in other embodiments the polynucleotide of interest is designed to inhibit the expression of two or more different soybean vacuolar processing enzymes. Thus in some embodiments, the soybean plants may be transformed with more than one polynucleotide of interest such as at least two polynucleotides of interest, at least three polynucleotides of interest, at least four polynucleotides of interest, at least five polynucleotides of interest, or at least six polynucleotides of interest, at least seven polynucleotides of interest, or more than seven polynucleotides of interest. When two or more polynucleotides of interest are transformed into the same plant cell, they may be expressed from the same expression cassette. Alternatively, the polynucleotides may be comprised in separate expression cassettes.

Various components of the expression constructs of the invention are described below.

A. Promoters

The promoter may be native or analogous or foreign or heterologous to the soybean plant host. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. When the promoter is "foreign" or "heterologous" to the plant host, it is intended that the promoter is not the native or naturally occurring promoter for the operably linked sequence encoding the polypeptide of interest

The nucleic acids can be combined with constitutive, tissue-preferred, or other promoters for expression in plants. Constitutive promoters include, for example, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO 99/43838 and U.S. Patent No. 6,072,050; the core CaMV 35S promoter (Odell *et al.* (1985) *Nature 313*:810-812); rice actin (McElroy *et al.* (1990) *Plant Cell 2*:163-171); ubiquitin (Christensen *et al.* (1989) *Plant Mol. Biol. 12*:619-632 and Christensen *et al.* (1992) *Plant Mol. Biol. 18*:675-689); pEMU (Last *et al.* (1991) *Theor. Appl. Genet. 81*:581-588); MAS (Velten *et al.* (1984) *EMBO J. 3*:2723-2730); ALS promoter (U.S. Patent No. 5,659,026), and the like. Other constitutive promoters include, for example, U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142.

Tissue-preferred promoters can be utilized to target enhanced expression of the polypeptide of interest within a particular plant tissue. Tissue-preferred promoters include Yamamoto *et al.* (1997) *Plant J. 12(2)*255-265; Kawamata *et al.* (1997) *Plant Cell Physiol. 38(7)*:792-803; Hansen *et al.* (1997) *Mol. Gen Genet. 254(3)*:337-343; Russell *et al.* (1997) *Transgenic Res. 6(2)*:157-168; Rinehart *et al.* (1996) *Plant Physiol. 112(3)*:1331-1341; Van Camp *et al.* (1996) *Plant Physiol. 112(2)*:525-535; Canevascini *et al.* (1996) *Plant Physiol. 112(2)*:513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol. 35(5)*:773-778; Lam (1994) *Results Probl. Cell Differ. 20*:181-196; Orozco *et al.* (1993) *Plant Mol Biol. 23(6)*:1129-1138; Matsuoka *et al.* (1993) *Proc Natl. Acad. Sci. USA 90(20)*:9586-9590; and Guevara-Garcia *et al.* (1993) *Plant J. 4(3)*:495-505. Such promoters can be modified, if necessary, for weak expression

"Seed-preferred" promoters include both "seed-specific" promoters (those promoters active during seed development such as promoters of seed storage proteins) as well as "seed-germinating" promoters (those promoters active during seed germination). See, Thompson *et al.* (1989) *BioEssays 10*:108, herein incorporated by reference. Such seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); milps (myo-inositol-1-phosphate synthase); and celA (cellulose synthase) (see the copending application entitled "Seed-Preferred Promoters," U.S. Application Serial No. 09/377,648, filed August 19, 1999, herein incorporated by reference). Gama-zein is a preferred endosperm-specific promoter. Glob-1 is a preferred embryo-specific promoter. For dicots, seed-specific promoters include, but are not limited to, bean β-phaseolin, napin, β-conglycinin (see, for example, Kitamura *et al.* (1984) *Theor. Appl. Genet.* 68:253-257, Cho *et al.* (1989) *Nucleic Acids Res.* 17:4386-4389, Kim *et al.* (1990) *Agric. Biol. Chem.* 54:1543-1550, Kim *et al.* (1990) *Protein Engineering* 3:725-731, Jung *et al.* (1998) *Plant Cell* 10:343-357, and Katsube *et al.* (1998) *BBA Gen. Subjects* 1379:107-117, herein incorporated by reference), soybean lectin, cruciferin, and the like.

B. Termination Regions

The termination region may be native with the transcriptional initiation region, may be native with the operably linked DNA sequence of interest, may be native with the plant host, or may be derived from another source (i.e., foreign or heterologous to the promoter, the DNA sequence of interest, the plant host, or any combination thereof). Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet. 262*:141-144; Proudfoot (1991) *Cell 64*:671-674; Sanfacon *et al.* (1991) *Genes Dev. 5*:141-149; Mogen *et al.* (1990) *Plant Cell 2*:1261-1272; Munroe *et al.* (1990) *Gene 91*:151-158; Ballas *et al.* (1989) *Nucleic Acids Res. 17*:7891-7903; and Joshi *et al.* (1987) *Nucleic Acid Res. 15*:9627-9639.

C. Leader Sequences

The expression cassettes may optionally contain 5' leader sequences in the expression cassette construct. Such leader sequences can act to enhance translation, for example, of a proteinase inhibitor polypeptide of the invention.. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein *et al.* (1989) *Proc. Natl. Acad. Sci. USA 86*:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Gallie *et al.* (1995) *Gene 165(2)*:233-238), MDMV leader (Maize Dwarf Mosaic Virus) (*Virology 154*:9-20), and human immunoglobulin heavy-chain binding protein (BiP) (Macejak *et al.* (1991) *Nature 353*:90-94); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling *et al.* (1987) *Nature 325*:622-625); tobacco mosaic virus leader (TMV) (Gallie *et al.* (1989) in *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256); and maize chlorotic mottle virus leader (MCMV) (Lommel *et al.* (1991) *Virology 81*:382-385). See also, Della-Cioppa *et al.* (1987) *Plant Physiol. 84*:965-968. Other methods known to enhance translation can also be utilized, for example, introns, and the like.

D. Selectable Marker Genes

Generally, the expression cassette will comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See generally, Yarranton (1992) *Curr. Opin. Biotech. 3*:506-511; Christopherson *et al.* (1992) *Proc. Natl. Acad. Sci. USA 89*:6314-6318; Yao *et al.* (1992) *Cell 71*:63-72; Reznikoff (1992) *Mol. Microbiol. 6*:2419-2422; Barkley *et al.* (1980) in *The Operon*, pp. 177-220; Hu *et al.* (1987) *Cell 48*:555-566; Brown *et al.* (1987) *Cell 49*:603-612; Figge *et al.* (1988) *Cell 52*:713-722; Deuschle *et al.* (1989) *Proc. Natl. Acad. Aci. USA 86*:5400-5404; Fuerst *et al.* (1989) *Proc. Natl. Acad. Sci. USA 86*:2549-

2553; Deuschle *et al.* (1990) *Science 248*:480-483; Gossen (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.* (1993) *Proc. Natl. Acad. Sci. USA 90*:1917-1921; Labow *et al.* (1990) *Mol. Cell. Biol. 10*:3343-3356; Zambretti *et al.* (1992) *Proc. Natl. Acad. Sci. USA 89*:3952-3956; Baim *et al.* (1991) *Proc. Natl. Acad. Sci. USA 88*:5072-5076; Wyborski *et al.* (1991) *Nucleic Acids Res. 19*:4647-4653; Hillenand-Wissman (1989) *Topics Mol. Struc. Biol. 10*:143-162; Degenkolb *et al.* (1991) *Antimicrob. Agents Chemother. 35*:1591-1595; Kleinschnidt *et al.* (1988) *Biochemistry 27*:1094-1104; Bonin (1993) Ph.D. Thesis, University of Heidelberg; Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. USA 89*:5547-5551; Oliva *et al.* (1992) *Antimicrob. Agents Chemother. 36*:913-919; Hlavka *et al.* (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.* (1988) *Nature 334*:721-724. Such disclosures are herein incorporated by reference.

The above list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be used in the present invention.

E. Polynucleotides of Interest

Because some of the soybean vacuolar processing enzymes of the invention have high levels of sequence identity in some regions, a polynucleotide of the invention may be designed to reduce or eliminate the activity of one or more vacuolar processing enzymes, for example, by targeting a region of the vacuolar processing enzyme mRNAs that are highly conserved. Alternatively, a polynucleotide may be designed to reduce or eliminate the activity of only one soybean vacuolar processing enzyme. Non-limiting examples of polynucleotides of interest are given below.

1. Sense Sequences

In some embodiments of the invention, inhibition of the expression of a vacuolar processing enzyme may be obtained by cosuppression. For cosuppression, the polynucleotide expressed by the expression constructs corresponds to all or part of an endogenous messenger RNA encoding a soybean vacuolar processing enzyme. The polynucleotide used for cosuppression may correspond to all or part of the messenger RNA encoding the vacuolar processing enzyme, all or part of the 5' and/or 3' untranslated

region of a vacuolar processing enzyme transcript, or all or part of both the coding sequence and the untranslated regions of a transcript encoding a vacuolar processing enzyme. In some embodiments where the polynucleotide comprises all or part of the coding region of the vacuolar processing enzyme, the expression cassette is designed to eliminate the start codon of the polynucleotide so that no protein product will be transcribed.

The sense sequence typically comprises at least 20 nucleotides, at least 50 nucleotides, at least 75 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 500 nucleotides, at least 1000 nucleotides, at least 5000 nucleotides, or more than 5000 nucleotides that correspond to a messenger RNA encoding a soybean vacuolar processing enzyme. The sense sequence generally has substantial sequence identity to the sequence of the transcript of the endogenous gene, preferably greater than about 65% sequence identity, more preferably greater than about 85% sequence identity, most preferably greater than about 95% sequence identity. See, U.S. Patent Nos. 5,283,184 and 5,034,323; herein incorporated by reference.

2. Antisense Sequences

In some embodiments of the invention, inhibition of the expression of a vacuolar processing enzyme may be obtained by antisense suppression. For antisense suppression, the expression cassette is designed to express nucleic molecule or interest corresponding to the complement of all or part of a messenger RNA encoding a soybean vacuolar processing enzyme . The polynucleotide for use in antisense suppression may correspond to all or part of the complement of the sequence encoding the vacuolar processing enzyme, all or part of the complement of the 5' and/or 3' untranslated region of a vacuolar processing enzyme transcript, or all or part of the complement of both the coding sequence and the untranslated regions of a transcript encoding a vacuolar processing enzyme.

Thus, antisense sequences are constructed to hybridize with the corresponding mRNA. Modifications of the antisense sequences may be made as long as the sequences hybridize to and interfere with expression of the corresponding mRNA. Thus, antisense

sequences may be fully or partially complementary to the target mRNA. In this manner, antisense constructions having 70%, preferably 80%, more preferably 85% sequence identity to the corresponding complements may be used. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, antisense sequences of at least 20 nucleotides, at least 50 nucleotides, at least 75 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 500 nucleotides, at least 1000 nucleotides, at least 5000 nucleotides, or more than 5000 nucleotides of the complement of the target mRNA may be used.

3. Polynucleotides for Double Stranded RNA Interference

In some embodiments of the invention, inhibition of the expression of a vacuolar processing enzyme may be obtained by double stranded RNA (dsRNA) interference. For dsRNA interference, a sense sequence like that described above for cosuppression and an antisense sequence that is complementary to the sense sequence are expressed in the same cell. The antisense sequence may be fully complementary to the sense sequence. Alternatively, the antisense sequence may be partially complementary to the sense sequence so long as it hybridizes to the sense sequence to form a double stranded RNA molecule.

Expression of the sense and antisense molecules can be accomplished by designing the expression cassette to comprise both a sense sequence and a complementary nucleotide sequence. Alternatively, separate expression cassettes may be used for the sense and complementary nucleotide sequences.

4. Polynucleotides for hpRNA Interference and ihpRNA Interference

In some embodiments of the invention, inhibition of the expression of one or more vacuolar processing enzyme may be obtained by hairpin RNA (hpRNA) interference or intron-containing hairpin RNA (ihpRNA) interference. For hpRNA interference, the expression cassette is designed to express nucleic molecule of interest that hybridizes with itself to form a hairpin structure that comprises a single-stranded

loop region and a base-paired stem. In some embodiments, the base-paired stem region is formed by hybridization between a sense sequence corresponding to all or a portion of a messenger RNA encoding a vacuolar processing enzyme and an antisense sequence that is complementary to the sense sequence. In other embodiments, the base-paired stem region is formed by hybridization between two sequences that are unrelated to an endogenous messenger RNA, and the loop region comprises all or part of the messenger RNA sequence for a soybean vacuolar processing enzyme.

Thus, in some embodiments, the sense sequence comprises at least 19, at least 30, at least 50, at least 100, at least 500, at least 1000, or more than 100 nucleotides corresponding to the mRNA encoding a soybean vacuolar processing enzyme (i.e. the target mRNA). The sense sequence generally shares at least 94% or more sequence identity with the corresponding region of the target mRNA, such as, for example, at least 95% or more sequence identity, at least 96% or more sequence identity, at least 97% or more sequence identity, at least 98% or more sequence identity, or at least 99% or more sequence identity. The antisense sequence may be fully complementary to the sense sequence. Alternatively, the antisense sequence may be partially complementary to the sense sequence so long as it hybridizes to the sense sequence to form a stem region. The hpRNA polynucleotide additional comprises a spacer or loop sequence operably 3' of the sense sequence and 5' of the antisense sequence. When the spacer sequence does not contain an intron, it is generally preferred to make the loop sequence as short as possible while still providing enough of a loop to allow the sense sequence to hybridize with the antisense sequence. Accordingly, the loop sequence is generally less than 1000 nucleotides, less than 900 nucleotides, less than 800 nucleotides, less than 700 nucleotides, less than 600 nucleotides, less than 500 nucleotides, less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, less than 100 nucleotides, or less than 50 nucleotides.

In other embodiments, the base paired stem structure is formed by the hybridization of a sense sequence that does not correspond to an endogenous sequence found in the host soybean plant, and an antisense sequence complementary to the sense sequence. The sense and antisense sequences flank a loop region that comprises all or

part of a sequence corresponding to a messenger RNA encoding a soybean vacuolar processing enzyme. Generally, the sense and antisense sequences will each be at least 40-50 nucleotides in length, such as 50-100 nucleotides in length, or 100-300 nucleotides in length. See, WO 0200904 for examples of sense and antisense sequences that may be used. The loop sequence corresponding to a messenger RNA encoding a soybean vacuolar processing enzyme generally comprises at least 25 nucleotides corresponding to the messenger RNA encoding the soybean vacuolar processing enzyme, and may comprise at least 50 nucleotides, at least 100 nucleotides, at least 200 nucleotides, or at least 300 nucleotides in length. The loop sequence generally shares at least 80% sequence identity with a messenger RNA encoding a soybean vacuolar processing enzyme, and may share at least 85% sequence identity, at least 90% sequence identity, or at least 95% sequence identity with a messenger RNA encoding a soybean vacuolar processing enzyme.

For ihpRNA, the interfering molecules have the same general structure as for hpRNA, but the RNA molecule additionally comprises an intron that is capable of being spliced in the cell in which the ihpRNA is expressed. The use of an intron minimizes the size of the loop in the hairpin RNA molecule following splicing, and this increase the efficiency of interference. Any intron that is spliced in soybean may be used according to the invention. Non-limiting examples of introns that may be used include the orthophosphate dikinase 2 intron 2 (pdk2 intron) described in U.S. Patent publication No. 20030180945, the catalase intron from Castor bean (Accession number AF274974), the Delta12 desaturase (Fad2) intron from cotton (Accession number AF331163), the Delta 12 desaturase (Fad2) intron from Arabidopsis (Accession number AC069473), the Ubiquitin intron from maize (Accession number S94464), and the actin intron from rice.

Transformation and Regeneration

In some embodiments, the methods of the invention comprise the steps of transforming and regenerating soybean plants. Suitable methods of introducing nucleotide sequences into plant cells and subsequent insertion into the plant genome include microinjection (Crossway *et al.* (1986) *Biotechniques 4*:320-334), electroporation

(Riggs *et al.* (1986) *Proc. Natl. Acad. Sci. USA 83*:5602-5606, *Agrobacterium*-mediated transformation (Townsend *et al.*, U.S. Patent No. 5,563,055; Zhao *et al.*, U.S. Patent No. 5,981,840), direct gene transfer (Paszkowski *et al.* (1984) *EMBO J. 3*:2717-2722), and ballistic particle acceleration (see, for example, Sanford *et al.*, U.S. Patent No. 4,945,050; Tomes *et al.*, U.S. Patent No. 5,879,918; Tomes *et al.*, U.S. Patent No. 5,886,244; Bidney *et al.*, U.S. Patent No. 5,932,782; Tomes *et al.* (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); and McCabe *et al.* (1988) *Biotechnology 6*:923-926) and Lec1 transformation (WO 00/28058). Also see Weissinger *et al.* (1988) *Ann. Rev. Genet. 22*:421-477; Sanford *et al.* (1987) *Particulate Science and Technology 5*:27-37 (onion); Christou *et al.* (1988) *Plant Physiol. 87*:671-674 (soybean); McCabe *et al.* (1988) *Bio/Technology 6*:923-926 (soybean); Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh *et al.* (1998) *Theor. Appl. Genet. 96*:319-324 (soybean); Datta *et al.* (1990) *Biotechnology 8*:736-740 (rice); Klein *et al.* (1988) *Proc. Natl. Acad. Sci. USA 85*:4305-4309 (maize); Klein *et al.* (1988) *Biotechnology 6*:559-563 (maize); Tomes, U.S. Patent No. 5,240,855; Buising *et al.*, U.S. Patent Nos. 5,322,783 and 5,324,646; Tomes *et al. (1995)* "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg (Springer-Verlag, Berlin) (maize); Klein *et al.* (1988) *Plant Physiol. 91*:440-444 (maize); Fromm *et al.* (1990) *Biotechnology 8*:833-839 (maize); Hooykaas-Van Slogteren *et al.* (1984) *Nature (London) 311*:763-764; Bowen *et al.*, U.S. Patent No. 5,736,369 (cereals); Bytebier *et al.* (1987) *Proc. Natl. Acad. Sci. USA 84*:5345-5349 (Liliaceae); De Wet *et al.* (1985) in *The Experimental Manipulation of Ovule Tissues,* ed. Chapman *et al.* (Longman, New York), pp. 197-209 (pollen); Kaeppler *et al.* (1990) *Plant Cell Reports 9*:415-418 and Kaeppler *et al.* (1992) *Theor. Appl. Genet. 84*:560-566 (whisker-mediated transformation); D'Halluin *et al.* (1992) *Plant Cell 4*:1495-1505 (electroporation); Li *et al.* (1993) *Plant Cell Reports 12*:250-255 and Christou and Ford (1995) *Annals of Botany 75*:407-413 (rice); Osjoda *et al.* (1996) *Nature Biotechnology*

14:745-750 (maize via *Agrobacterium tumefaciens*); all of which are herein incorporated by reference.

The cells that have been transformed may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having constitutive expression of the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved.

PLANTS AND SEED

The invention also provides soybean plants that are genetically modified or mutagenized to reduce or eliminate the activity of one or more vacuolar processing enzymes in seed, and transformed seed of these plants. The term "genetically modified" as used herein refers to a plant cell or plant that is modified in its genetic information by the introduction of one or more foreign polynucleotides, and that the expression of the foreign polynucleotides leads to a phenotypic change in the plant. By "phenotypic change," it is intended a measurable change in one or more cell functions. For example, the genetically modified plants of the present invention show reduced or eliminated expression or enzymatic activity of one or more vacuolar processing enzymes. Also provided are soybean plants that have been mutagenized and carry a mutation in one or more genes encoding a vacuolar processing enzyme that results in reduced activity of the encoded vacuolar processing enzyme.

The soybean plants encompassed by the invention may be genetically modified or mutated to inhibit the expression or enzymatic activity of at least one, at least two, at least three, at least four, at least five, at least six, or at least seven or more vacuolar processing enzymes. Those of ordinary skill in the art recognize that this can be accomplished in any one of a number of ways. For example, each of the expression cassettes for inhibiting the expression or enzymatic activity of the vacuolar processing enzymes can be

operably linked to a promoter and then joined together in a single continuous fragment of DNA comprising an expression cassette. Such an expression cassette can be used to transform a plant to produce the desired outcome. Alternatively, separate plants can be transformed with expression cassettes capable of expressing a polynucleotide, which inhibits the expression of different vacuolar processing enzyme. A single plant that is genetically modified to inhibit the expression or the enzymatic activity of two or more vacuolar processing enzymes can then be produced by transforming a selected genetically modified plant to inhibit the expression of a different vacuolar processing enzyme, and selecting for plants showing inhibition in expression or enzymatic activity of multiple vacuolar processing enzymes. Multiple rounds of transformation and selection may be required to produce the desired plant.

Alternatively, a single plant that is genetically modified or mutagenized to inhibit the expression or the enzymatic activity of two or more vacuolar processing enzymes can be produced through one or more rounds of cross pollination utilizing the previously selected seed-protease deficient plants as parents. Methods for cross pollinating plants are well known to those skilled in the art, and are generally accomplished by allowing the pollen of one plant, the pollen donor, to pollinate a flower of a second plant, the pollen recipient, and then allowing the fertilized eggs in the pollinated flower to mature into seeds. Progeny containing the entire complement of heterologous coding sequences of the two parental plants can be selected from all of the progeny by standard methods available in the art as described *supra* for selecting transformed plants. If necessary, the selected progeny can be used as either the pollen donor or pollen recipient in a subsequent cross pollination.

METHODS OF DETERMINING % SEQUENCE IDENTITY

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988) *CABIOS 4*:11-17; the local homology algorithm of Smith *et al.* (1981) *Adv. Appl. Math. 2*:482; the

homology alignment algorithm of Needleman *and Wunsch* (1970) *J. Mol. Biol. 48*:443-453; the search-for-similarity-method of Pearson *and Lipman* (1988) *Proc. Natl. Acad. Sci. 85*:2444-2448; the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA 872*64, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA 90*:5873-5877.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 10 (available from Accelrys Inc., 9685 Scranton Road, San Diego, California USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (1988) *Gene 73*:237-244 (1988); Higgins *et al.* (1989) *CABIOS 5*:151-153; Corpet *et al.* (1988) *Nucleic Acids Res. 16*:10881-90; Huang *et al.* (1992) *CABIOS 8*:155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol. 24*:307-331. The ALIGN program is based on the algorithm of Myers and Miller (1988) *supra*. A PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used with the ALIGN program when comparing amino acid sequences. The BLAST programs of Altschul *et al* (1990) *J. Mol. Biol. 215*:403 are based on the algorithm of Karlin and Altschul (1990) *supra*. BLAST nucleotide searches can be performed with the BLASTN program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the invention. BLAST protein searches can be performed with the BLASTX program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to a polypeptide of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res. 25*:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.* (1997) *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins)

can be used. See www.ncbi.hlm.nih.gov. Alignment may also be performed manually by inspection.

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 using the following parameters: % identity using GAP Weight of 50 and Length Weight of 3; % similarity using Gap Weight of 12 and Length Weight of 4, or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

GAP uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443-453, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the

number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

EXPERIMENTAL

Altered Solubility Profile for *Arabidopsis thaliana* Seed Storage Proteins in the Absence of Vacuolar Processing Enzyme Activity

I. Methods

A. Isolation of the *αvpe::dSpm1* Allele

A putative *dSpm* transposon insertion in *αVPE* was identified in DNA of SLAT (Sainsbury Laboratory *Arabidopsis thaliana dSpm* Transposants) pool 5.38 by probing a filter blot, obtained from the Sainsbury Laboratory displaying flanking DNA of the Sainsbury *dSpm* transposon insertion population, with a genomic DNA probe corresponding to the *αVPE* gene.

Confirmation and localization of the *dSpm* insertion within *αVPE* (*αvpe::dSpm1* allele) was accomplished by PCR of pool 5.38 genomic DNA (obtained from the Sainsbury laboratory), PCR product isolation, and DNA sequencing as previously described. Plants homozygous for the *αvpe::dSpm 1* allele were identified by PCR from progeny of the 5.38 seed pool. Homozygosity was confirmed by the lack of PCR detectable wild-type alleles in the F2 progeny following self-pollination of putative *αvpe::dSpm 1* homozygous plants.

B. Isolation of The *γvpe::T-DNA1* Allele

The SIGnAL (Salk Institute Genomic Analysis Laboratory) database (available at signal.salk.edu/cgi-bin/tdnaexpress) of T-DNA left border adjacent sequences was queried with the *γVPE* sequence to identify a seed stock (Salk_010372) containing an insertion within the 5th exon of *γVPE*. Seeds from this line were obtained from the Arabidopsis Biological Resource Center (ABRC), and seedlings screened by PCR to identify plants homozygous for the *γvpe::T-DNA 1* allele. DNA was isolated with the DNeasy Plant Mini kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's protocol and subjected to PCR to detect the *γvpe::T-DNA 1* allele. Homozygous *γvpe::T-DNA* plants were confirmed by the lack of PCR detectable wild-type alleles in the F2 progeny following self-pollination.

C. Genetic Stacking and PCR Identification of Homozygous Mutants

Genetic stacking and isolation of VPE mutant plants was performed as follows. First, plants homozygous for both the *βvpe::dSpm1* and *δvpe::dSpm1* alleles (Gruis *et al.* (2002) *Plant Cell* 14:2863-82) were crossed with plants homozygous for *αvpe::dSpm1*. Second, plants among the segregating F2 progeny (following F1 self pollination) identified as homozygous for *αvpe::dSpm1*, *βvpe::dSpm1* and *εvpe::dSpm1* were then crossed with plants homozygous for *γvpe::T-DNA1*. For PCR screening of F2 progeny following F1 self pollination of the second cross, DNA was prepared from one rosette leaf of each plant prior to flowering. Fresh tissue was harvested into 1.1 ml minitubes of a 96-well Megatiter-Plate (Biological Band Continental Lab Products) on ice. A 5/32" steel bead and 200 µl of extraction buffer (10% w/v potassium ethyl xanthogenate, 100 mM Tris pH 7.5, 2 M NaCl and 10 mM EDTA) were added to each sample immediately prior to homogenization in a Raptor/Geno/Grinder (Spex CentiPrep Inc., Metuchen, NJ) for 1 minute at 7000 strokes/minute. Following incubation at 65°C for 30 minutes, the samples were cooled on ice for 15 minutes, centrifuged at 3,000*g* for 15 minutes and 150 µl of supernatant transferred to a new tube. A second centrifugation was again performed to remove debris and 100 µl of supernatant was transferred to a new tube containing 150

µl of ice cold 2-propanol. The DNA-precipitate was pelleted by centrifugation, rinsed with 300 µl of cold 70% ethanol v/v, dried for 20 minutes in a 65°C air incubator and incubated at 65°C for 20 minutes with 150 µl of 5 mM Tris-HCl pH 8.0. 3 µl of this DNA preparation were used per PCR reaction. The putative genotypes of selected plants of interest identified from the initial large scale screen were then confirmed by a second round of PCR analysis using DNA isolated from an independently harvested rosette leaf with the DNeasy Plant Mini kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's protocol. Homozygosity of the various mutant allele combinations was confirmed by the lack of detectable wild-type alleles in the F3 progeny following self-pollination.

D. *γVPE* Knock-Down/*βvpe* Plants

Confirmation of the *γVPE* null-allele phenotype was accomplished by transforming *βvpe* mutant plants with an intron-spliced self-complimentary hairpin RNAi construct (Smith *et al.* (2000) *Nature* 407:319-320) designed to knock down *γVPE* expression. The RNAi portion of the vector was constructed using standard cloning techniques to splice the *β* phaseolin promoter described by Slightom *et al.*, Custom polymerase chain reaction engineering plant expression vectors and genes for plant expression, pp. 1-55 in *Plant Molecular Biology Manual*, Gelvin and Schilperoort, eds., Dordrect:Kluwer Academic Publishers (1991), with an rtPCR-amplified 500 bp fragment (nucleotides 27-526 of NCBI Accession No. AF370160) of *γVPE* in the sense orientation, a 1133 bp PCR-amplified FAD2 intron sequence (nucleotides 142-1274 of NCBI Accession No. AC069473), and a 500 bp fragment of *γVPE* in reverse orientation. The transformation vector also contained the constitutive promoter SCP1 described by U.S. Patent No. 6,555,673 to Bowen *et al.* to drive expression of the selectable marker, the neomycin phosphotransferase II gene. *Agrobacterium*-mediated transformation using strain GV3101 carrying the helper plasmid pMP90 was performed using the flora dip method described by Clough and Bent (1998) *Plant J.* 16:735-43). Kanamycin resistant seedlings were selected, allowed to self-pollinate, and T1 seed of *γVPE* knock-down events were analyzed by SDS-PAGE.

E. SDS-PAGE and Immunoblotting

Developing, germinating and mature seed were collected and protein was extracted under reducing conditions as described by Gruis *et al.* (2002) *Plant Cell* 14:2863-82. Protein extraction for SDS-PAGE under oxidizing conditions was accomplished by homogenization of mature seed meal with a 20-fold v/w excess of ice-cold 2% SDS, 50 mM Tris-HCl pH 6.8, and 100 mM iodoacetamide. Samples were incubated on ice, for 5 minutes at room temperature, and finally for 5 minutes at 100° C. After incubation, the samples were treated as reduced protein extracts as described in Gruis *et al.* (2002) *Plant Cell* 14:2863-82, except that DTT was omitted from SDS-PAGE sample buffer. Proteins were electrophoretically separated by SDS-PAGE using one of the following methods: Tris-Tricine gels (8% spacer and 15% separating), Tris-Tricine gels using a 8% spacer and a 12% separating gel or Tris-Glycine 4-20% gradient mini-gels (BioRad, Hercules, CA). Immunoblotting was performed using either a 1:2500 dilution of anti-sera generated using rape seed cruciferin to detect legumin-type globulins or a 1:5000 dilution of anti-sera generated using HPLC-purified Arabidopsis napin-type albumins. The legumin-type globulin anti-sera cross reacts with α-chain epitopes of Arabidopsis legumin-type globulins and the Arabidopsis napin-type albumin specifically detects epitopes on the large chains.

F. Linear Sucrose Density Gradient Separations

Dry mature seed was ground at room temperature using a porcelain mortar and pestle and 25 mg of the resulting meal was defatted in 2 ml microcentrifuge tubes by three sequential 1 ml hexane extractions at room temperature. Following vacuum desiccation, the meal was re-suspended in 20 v/w ice cold extraction buffer (100mM sodium phosphate buffer pH 7, 400 mM KCl) containing 1mM Pefabloc (Roche Molecular Biochemicals, Indianapolis, IN) and incubated at 4°C for 40 minutes with constant agitation. The supernatant was then recovered following a 10 min centrifugation at 20,800*g* and the protein concentration was determined using the bicinchoninic acid (BCA) Protein Quantitation Assay (Pierce, Rockford, IL) standardized using bovine

serum albumin (Pierce, Rockford, IL). Following extraction, protein samples were immediately loaded onto sucrose density gradients.

Linear sucrose density gradients (6-20%) were prepared in SW40 ultracentrifuge tubes (Beckman Coulter Instruments Inc., Fullerton, CA) using the BIOCOMP Gradient Maker 107ip (BioComp Instruments Inc., New Brunswick, Canada) per the manufacturer's instructions. 200 µl of protein extract (~1.5 mg of protein) was applied to the top of the prepared gradients. Proteins were then fractionated by centrifugation at 37,000 rpm (SW40 rotor) at 4°C for 21 hours. Following centrifugation, gradients were fractionated using a BIOCOMP Piston Gradient Fractionator-151 (BioComp Instruments Inc., New Brunswick, Canada) at 0.3 mm/sec and collected using a Frac-200 fraction collector (Pharmacia LKB, Uppsala, Sweden) set up to collect 12 drops (~300 µl) per fraction. Any potential pellet remaining at the bottom of the tube was re-suspended in 100 µl of SDS protein extraction buffer for analysis. The protein quantity in each gradient fraction was determined using the BCA assay (Pierce, Rockford, IL) and results plotted for each fraction as a percentage of the protein detected in all fractions. Proteins of known sedimentation coefficients; chymotrypsin (2.6S), bovine serum albumin (4.4S), aldolase (7.3S) and catalase (11.3S) (Pharmacia LKB, Uppsala, Sweden) were separated in parallel gradients and used as a reference to assign sedimentation coefficients to the Arabidopsis seed protein gradient fractions.

Prior to analysis by SDS-PAGE each gradient fraction sample was concentrated 5 fold using Micron YM-3 centrifugal filter devices (Millipore, Bedford, MA). For Coomassie Brilliant Blue R-250 stained SDS-PAGE analysis, 10 µl of sample was incubated at 100°C for 5 minutes with 4 µl of SDS-PAGE loading buffer (250 mM Tris pH 6.8, 500 mM DTT, 10% w/v SDS, 0.5% w/v bromophenol blue, 50% v/v glycerol). Samples were then electrophoresed in 26-well 4-20% gradient Tris-HCl mini-gels (BioRad, Hercules, CA). Immunoblotting was carried out as described using 2 µl of each sample.

G. Solubility Profiling

Proteins were extracted and separated using linear sucrose density gradients (see above). Legumin-type globulin protein from wild-type seed was obtained by pooling fractions #24-30 from 4 parallel linear sucrose density gradient separations of wilt-type seed proteins. Legumin-type globulin protein from *vpe*-quad mutant seed was obtained by pooling fractions #15-21 from 4 parallel linear sucrose density gradients of *vpe*-quad seed proteins. Proteins contained in these pooled fractions were first subjected to a 1500 fold dilution into buffer (150 mM NaCl, 20 mM Tris pH 8.0) and subsequently concentrated to ~20 mg/ml using Amicon Ultra 10,100 MWCO centrifugal filter devices (Millipore, Bedford, MA) according to manufacturer's instructions. Following this procedure each protein sample was quantified and adjusted to a final concentration of 14 mg/ml using the BCA assay (Pierce, Rockford, IL). For each sample (12S wild-type and 9S quad) dilutions of protein into several pH buffers (Na Acetate-acetic acid, pH 3.5, pH 4.0, pH 4.5, pH 5.5; MES-NaOH pH 5.5, pH 6.0, pH 6.5; Hepes-HCl, pH 7.0, pH 7.5, pH 8.0; Tris-HCl pH 8.5) was performed at room temperature. Each pH condition was set up as a 30 µl reaction mixture in a microcentrifuge tube containing a final concentration of 25 mM buffer, 10 mM NaCl and 0.9 mg/ml protein. Following incubation at room temperature for 2 hours, samples were subjected to centrifugation at 20,800*g* for 10 min. Supernatants were then assayed for protein content using the BCA assay (Pierce, Rockford, IL) and results for each sample plotted as a percentage of protein remaining in the supernatant (soluble).

II. Results

A. Detection of Vegetative-Type VPE Gene Expression in Developing Seed

Because vegetative-type VPE gene expression is induced in vegetative-tissues under stress conditions (Kinoshita *et al.* (1999) *Plant J.* 19:43-53), the possibility that vegetative-type VPE gene expression may be induced due to abnormal accumulation of precursor proteins in βVPE mutant seed was tested. Semi-quantitative multiplexed RT-PCR was performed using γVPE specific primers in combination with primers specific

for a constitutively expressed transcript (cytosolic ribosomal protein S11). This analysis detected γVPE transcript in a vegetative control sample (leaf), known to express γVPE. However, contrary to expectations, prominent γVPE-specific amplification products were also detected in developing seed of wild-type plants. The ratio of the intensity of the γVPE-specific band compared to the S11- specific band indicated similar amounts of γVPE transcript were present in leaf and developing seed samples of wild-type and βVPE/εVPE double mutants. To confirm and quantify γVPE transcript in developing seed, quantitative real-time PCR was performed using independently isolated RNA from developing seed of both wild-type and the βvpe/εvpe double mutants. This analysis also detected γVPE transcript in developing wild-type seed and showed no significant change of γVPE transcript level in the mutant sample.

To further substantiate this observation and to relate the quantity and/or significance of γ VPE expression in seed to the other members of the VPE gene family, queries of several Arabidopsis Massively Parallel Signature Sequencing (MPSS) high-resolution gene expression datasets with conceptual MPSS expressed sequence tags (ESTs) of Arabidopsis VPE genes were performed. MPSS gene expression datasets are essentially EST sequencing experiments each consisting of 1 to 2 million independently derived MPSS ESTs from a single tissue source. Therefore, these very deep EST sequence libraries provide quantitative gene expression data reported in parts per million (ppm) for each transcript. Corroborating the RT-PCR results, γVPE transcripts are present in developing seed concurrently with βVPE and δVPE transcripts. Moreover, the second Arabidopsis vegetative-type VPE gene, αVPE, is also expressed in developing seed, albeit at much lower levels (4-10-fold less) than γVPE. The βVPE expression profile is similar to the expression profile of seed storage protein genes, showing peak expression in seed 14 days after anthesis. At this stage, βVPE is the most prominent VPE gene transcript detected, approximately 3-fold more prevalent than γVPE transcript. γVPE transcript is the second most abundant VPE gene transcript detected at this stage (MSS), however, 2-3 fold higher levels of this transcript are detected earlier during seed development. γVPE is also the only VPE gene for which significant levels of transcript

are detected in vegetative tissues including leaves and roots. The δVPE gene is the most abundant VPE gene transcript during the cell division stage of seed development and in germinating seed. δVPE transcript is also present at significant levels in all other developing seed stages assayed. Together, these data indicate that all four Arabidopsis VPE genes, including vegetative-type VPE family members, are significantly expressed in developing seed during storage protein accumulation.

B. Isolation of Vegetative-type VPE Gene Knock-Out Mutants

To investigate a potential function of the two Arabidopsis vegetative-type VPE genes during seed development, plants containing DNA insertion alleles in the αVPE and γVPE genes were isolated. A putative dSpm transposon insertion allele of αVPE (αvpe::dSpm1) was identified in pool 5.38 of the Sainsbury Laboratory collection by reverse screening using SLAT blots probed with DNA of α-VPE. DNA flanking the insertion site was cloned and sequenced to determine the location of the dSpm element within the gene. The dSpm insertion in αvpe::dSpm1 is located 249 bp downstream of the translational start codon in the intron following the first exon of the gene. The dSpm element used in creating the Sainsbury mutant collection has been designed to contain transcriptional stop sites in either orientation such that intronic insertion events would interfere with gene transcription. To test whether αvpe::dSpm1 is a knock-out allele, multiplexed RT-PCR using αVPE-specific primers annealing downstream of the dSpm insertion site in combination with primers specific for a control transcript (cytosolic ribosomal protein S11) was performed with RNA isolated from 14 DAA seed of two homozygous αvpe::dSpm1 plants and from two wild-type plants. A PCR product corresponding to αVPE transcript was amplified only in wild-type seed samples and not in samples of seed homozygous for the αvpe::dSpm1 allele, classifying the αvpe::dSpm1 allele as a null-allele.

A putative T-DNA insertion allele of γVPE (γvpe::T-DNA1) was identified by querying the SIGnAL website (available at salk.edu). Seed from the corresponding mutant line (Salk_010372) was obtained from the Arabidopsis Biological Resource

Center and plants homozygous for the γvpe::T-DNA1 allele were subsequently identified using allele specific PCR. Analysis of the T-DNA adjacent DNA sequence was used to identify the T-DNA integration site as located within exon 5 of the γVPE gene. To test whether γvpe::T-DNA1 is a null allele, RT-PCR was performed essentially as described above for αvpe::dSpm1. γVPE transcript was clearly detected in wild-type control plants but not in homozygous γvpe::T-DNA1 plants, a result indicative of a knock-out allele.

Mutants homozygous for either αvpe::dSpm1 or γvpe::T-DNA1 were examined for visible phenotypes under normal growth conditions. No effects were observed on germination rate, vegetative growth rate, plant architecture, seed set, or senescence compared to wild-type controls. Moreover, no differences between protein profiles of mutant and wild-type seed were detected.

C. Genetic Stacking of VPE Mutant Alleles

Genetic stacking of null-alleles of the four unlinked Arabidopsis VPE genes was performed. A βvpe/δvpe double mutant was first crossed to the αvpe mutant and triple mutant plants (αvpe/βvpe/δvpe), homozygous for the respective null-alleles at each locus, were identified by allele-specific PCR analysis of the segregating F2 progeny following F1 self-pollination. The αvpe/βvpe/δvpe triple mutant was then crossed to the γvpe mutant and, after F1 self-pollination, a total of 1132 F2 progeny plants were screened for the absence and presence of wild-type and mutant alleles at each VPE locus. This screen identified two αvpe/βvpe /γvpe/δvpe quadruple-mutant plants (referred to herein as *vpe*-quad) homozygous for null-alleles at all four VPE loci, as well as plants with all possible combinations of homozygous triple-mutant alleles and homozygous double mutant alleles of VPE genes. A minimum of two plants of each genotype was isolated (not all data shown). Progeny of these plants, including *vpe*-quad plants, were grown for two generations under normal growth conditions side-by-side with wild-type plants and closely inspected for any phenotypic variation compared to the wild-type controls. In all cases, no effects were observed on germination rate, vegetative growth, flowering time, seed set, senescence, plant architecture or light-microscopic seed morphology.

D. Seed Protein Profiles of VPE Mutants

The impact of removal of VPE expression on seed storage protein processing was examined with seed protein extracts (Figure 1) from plants with the mutant allele combinations described in the description of the figure. A minimum of two plants of each genotype were analyzed to ensure that SDS-PAGE protein profiles shown in Figure 1 are representative for each investigated genotype. Several observations can be made from this gel analysis. The double null-mutant of the vegetative-type VPE genes (αvpe /γvpe) does not detectably alter seed protein processing. Mutants of seed-type VPEs, either βvpe or βvpe /δvpe double mutants, show subtle changes in the mature seed protein profiles. The combination of the βvpe /δvpe double mutants with the vegetative-type αvpe mutant (αvpe/βvpe/δvpe) do not result in any discernable additional change in the protein profile beyond what is observed for the seed-type VPE mutants alone. However, dramatic differences in protein profiles are observed in seeds of plants that are homozygous for null-alleles at both the βVPE loci and γVPE loci. The accumulation of polypeptides of the apparent molecular mass predicted for pro-protein forms of the legumin-type globulin proteins is increased while polypeptides corresponding to mature α- and β- chains are significantly decreased. Additionally, accumulation of the mature small chains of napin-type albumins is decreased and polypeptides of apparent molecular mass greater than that observed for mature large chains significantly accumulate. Interestingly, the comparison of the protein of the βvpe/γvpe/δvpe mutants with the protein profile of *vpe*-quad mutants reveals subtle additional changes of legumin-type globulin and napin-type albumin accumulation that can be attributed to the αvpe null-allele. Therefore, both vegetative-type VPEs are involved in seed protein processing.

To independently corroborate the observed null-allele phenotype of vegetative-type VPEs, a βvpe mutant plant was transformed with a RNA silencing construct to suppress γVPE expression. The seed protein profile from a resulting γVPE knock-down/βvpe plant is similar to that observed for βvpe/γvpe/εvpe triple mutants supporting

the conclusion that the observed seed protein profile phenotypes of the vegetative-type VPE mutants are indeed a direct result of the insertional interruption of VPE genes.

E. Alternative Proteolytic Processing of Seed Proteins

In addition to detecting polypeptides of an apparent molecular mass consistent with pro-forms of legumin-type globulins, several novel polypeptides of lesser molecular masses were observed in *vpe*-quad seed under reducing SDS-PAGE conditions. At least some of these polypeptides cross-reacted with α-chain specific legumin antibodies identifying them as alternatively processed legumin-type globulin polypeptides containing α-chain epitopes. To determine if any of the other novel polypeptides are disulfide-linked to these legumin α-chain-related polypeptides, seed proteins were extracted in the presence of iodoacetamide (IAA) and separated by SDS-PAGE under oxidizing conditions. Alkylation of free sulfhydryl groups with IAA was necessary to prevent disulfide interchange reactions in legumin-type globulin subunits. Without IAA added, even under oxidizing conditions, these reactions caused extensive breakage of disulfide-bonds between α- and β-chains of Arabidopsis legumin-type globulins. As expected, under oxidizing SDS-PAGE conditions, wild-type seed protein bands shifted to apparent molecular masses consistent with legumin-type pro-globulins (~ 50kD) and napin-type pro-albumins (~12kD), indicative of disulfide linked chains for each class of storage proteins. When IAA-treated protein from the *vpe*-quad seed was analyzed, it was likewise evident that many of the novel polypeptides observed under reducing SDS-PAGE conditions were size-shifted under oxidizing conditions. Most polypeptides appeared to migrate at sizes similar to pro-proteins, including the bands that corresponded to legumin-type globulin polypeptides with α-chain epitopes. However, at least one of these legumin-specific bands (~40kD) appears to be smaller than legumin-type pro-globulins, indicating alternative cleavage that results in the loss of a polypeptide chain (~10kD), which is not disulfide-linked to the alternatively processed subunit. Additionally in *vpe*-quad seed, napin-type albumins, size shifted under oxidizing conditions, are slightly greater in apparent molecular mass than the napin-type polypeptides accumulated in wild-type. This observation is consistent with efficient

VPE-independent cleavage of napin-type pro-polypeptides into disulfide linked large and small chains that contain additional amino acids.

F. N-terminal Amino Acid Sequence Analysis

To further investigate the nature of alternative processing in developing *vpe*-quad seed, Edman degradation was performed for several prominent polypeptide bands that appeared to be novel compared to wild-type. Separation of seed proteins using linear sucrose density gradients and SDS-PAGE was used to further enrich protein bands prior to sequencing. All polypeptides successfully identified from the *vpe*-quad 9S and 2S fractions were derivatives of legumin-type globulins and napin-type albumins respectively. The majority of identifications corresponded to the two most highly expressed seed storage protein genes, legumin-type globulin cruciferin 1 and napin-type albumin 3.

Six polypeptides were successfully sequenced and identified from the 9S fraction of *vpe*-quad. The N-terminal sequence of two polypeptides with an apparent molecular mass consistent with pro-forms of legumin-type globulins, each corresponded to the sequence of a different legumin-type globulin immediately downstream of the predicted signal peptide. Therefore, sequence and molecular mass identify these two legumin-type globulin proteins as unprocessed precursors.

Instead of mature β-chains of legumin-type globulins, *vpe*-quad seed accumulated prominent polypeptides that are approximately 1kD greater in molecular mass than β-chains accumulated in wild-type seed. Similar to wild-type β-chains, these proteins failed to bind α-chain specific legumin anti-sera. The N-terminal sequence obtained for one of these polypeptides corresponded to the hyper-variable region sequence of a legumin-type globulin, 11 residues upstream of the Asn-Gly polypeptide bond that is normally cleaved in wild-type seed by VPE. A second polypeptide matched the N-terminal sequence immediately downstream of the signal peptide. However, the apparent mass of this polypeptide was ~ 32 kD, which is 1-2 kD less than the calculated mass for the mature α-chain derived from this protein. The sizes and sequences of the polypeptides with band ID 6 and 10 are therefore consistent with the same alternative

cleavage event occurring in the hyper-variable region of the legumin-type globulin, upstream of the normally processed Asn-Gly bond.

In addition to proteolytic cleavage of legumin-type globulins yielding novel α– and β– chain-like fragments, other fragments of lesser molecular mass than either α- or β- chains were also identified. Several polypeptides that were all derived from a single legumin-type globulin gene were identified, indicating that no single preferred alternative-processing pathway appeared to exist to compensate for the lack of VPE activity. N-terminal amino acid sequencing of napin-type albumin polypeptides isolated from *vpe*-quad seed allowed for the successful identification of most of these polypeptides. The vast majority of napin-type albumin did not accumulate as a precursor-like form, but is instead processed to novel forms.

All cleavage sites of napin-type albumins so far identified by amino-terminal sequencing in *vpe*-quad seed involved a Phe residue at the P1 or P1' position. Additionally, the cleavage of at least one legumin-type polypeptide also occurred at a Phe in P1'. Proteolysis at these locations is consistent in sequence context with cleavage by a member(s) of the aspartic protease gene family.

G Impact of Processing on Legumin-type Globulin Solubility

The solubility profile of legumin-type globulins changes following VPE-specific processing of pro-forms into mature α- and β- chains such that a profound decrease in solubility under acidic conditions (pH 4.5-5.5) is observed. To determine if legumin-type globulin accumulated in *vpe*-quad seed shares similar solubility properties with wild-type VPE-processed protein, the solubility profile of the wild-type 12S proteins was compared to the 9S proteins of *vpe*-quad (Figure 2). The solubility profile of VPE-processed legumin-type globulin (wild-type) shows the protein to be largely soluble at pH 7-8.5 and 3.5-4. At intermediate pH ranges, the solubility of the wild-type protein fraction is gradually reduced with the majority of protein being insoluble at pH 5.5-6.0. Contrasting this result, the solubility profile of legumin-type globulin accumulated in *vpe*-quad seed shows the protein to be mostly soluble at pH 7.5-8.5, and mostly insoluble at pH 3.5-5. See Figure 3. The solubility of the protein at intermediate pH 5.5-6.0 is

~60-70%. Therefore the solubility profile of the legumin-type globulin accumulated in *vpe*-quad seed is markedly altered compared to wild-type supporting a function of proteolytic processing in determining this physiochemical property.

III. Conclusions

A. Vegetative-type VPE Expression in Developing Seed

A common theme of storage protein deposition in the PSV of plant seeds is pro-protein processing by proteolytic cleavage at Asn residues in the P1 position of cleavage sites. Prior to the present disclosure, vegetative-type VPE genes were not believed to be involved in Asn-specific storage protein processing because earlier studies strongly implied that vegetative-type VPE genes encode isoforms of VPE that are not expressed in seed, but are specific to vegetative tissues. The RT-PCR detection of significant amounts of γVPE message in developing seed of wild-type plants was therefore a surprising result. However, this result is firmly supported by the MPSS transcript profiles obtained for the VPE genes. Although the MPSS analysis corroborated prior reports of γVPE expression in leaf and βVPE expression in developing seed, it also clearly showed that expression of these VPE genes are not mutually exclusive to those tissues as previously implied. The present analysis identified expression of all four VPE genes in developing seed, with transcript levels of each VPE gene exceeding those measured in non-seed tissues (root, leaf, shoot inflorescences).

B. Functions of VPE genes

Interestingly, the expression patterns of the VPE genes appear to be significantly different from each other, yet at least three of the four genes in Arabidopsis seem to be involved in seed storage protein processing. It may expected that VPE gene functions are difficult to identify in many cases from single or even double mutants as overlapping or induced expression will act in a compensatory fashion similar to what we observed with single gene VPE mutants in seed protein processing. However, this would not be expected to occur in the *vpe*-quad mutant for which all VPE genes identified in the Arabidopsis genome are knocked out, and in fact is confirmed by examination of seed

protein processing in this report. Surprisingly, despite VPE being implicated in several processes throughout plant growth and development, no deleterious or pleiotropic effects of not having a functional VPE protease were detected.

C. Seed Proteins are Processed by Vegetative-type VPE

To measure the specific contribution of αVPE and γVPE to storage protein processing it was necessary to obtain seed from plants homozygous for additional combinations of VPE mutant alleles. Investigation of the seed protein profiles from either βvpe/γvpe or αvpe/βvpe/γvpe clearly identified increased accumulation of legumin-type globulin precursors indicating that both seed- and vegetative-type VPE can perform roles in storage protein processing. Additionally, no wild-type α- or β- chains of legumin-type globulins could be identified in seed devoid of αVPE, βVPE and γVPE supporting the hypothesis that VPEs are unique in their responsibility to process legumin-type globulin storage proteins at the conserved Asn-Gly peptide bond separating the chains. Furthermore, this exclusive responsibility extends to Asn-specific napin-type albumin processing as no wild-type small chains were found in *vpe*-quad. Also, similar to what was reported for βVPE, no evidence linking a specific VPE gene to proteolytic processing of a specific subset of legumin-type or napin-type storage proteins was found. Therefore, both the *in planta* functional analysis of VPE mutant Arabidopsis plants and the VPE gene expression analysis does not support the paradigm of two strict VPE classes, seed-type and vegetative-type, performing entirely separate functions as previously proposed. Instead, evidence presented here suggests that VPE gene family members have multiple expression patterns, and overlapping functions in at least developing seed.

D. Processing and Storage Protein Accumulation Mechanisms

Mature VPE-processed legumin-type globulin from soybean (glycinin) is considerable less soluble under acidic conditions at pH 4-6 when compared to bacterially expressed precursors of glycinin. VPE-processed Arabidopsis legumin-type globulins are also mostly insoluble at pH 5.5-6, which coincides with the pH of the PSV in developing

seed. Although, alternatively processed legumin-type globulins in *vpe*-quad appear to be partial soluble at pH 5.5, they are insoluble under more acidic conditions. These data show that the specific solubility properties are impacted by the processing status of legumin-type globulin polypeptides. Recently it has been shown that an intermediate form of a drought responsive cysteine protease (iRD21) is insoluble under acidic conditions and is forming aggregates in vacuoles. Further, it has been suggested, that this aggregate may functions as a stock of inactive protease that could be made soluble under the appropriate physiological conditions to be available as an active enzyme. Similar to iRD21, aggregation of globulins in PSV, perhaps induced by limited proteolytic processing, could serve as a mechanism to ensure long-term stable globulin storage by sequestering these proteins away from the lytic conditions of the vacuole. During germination, storage proteins could be mobilized from these aggregates by a change of the pH or of the ionic strength of the vacuole, which would render the proteins soluble and make them accessible to proteolytic enzymes.

Inhibition of the Expression of Vacuolar Processing Enzymes in Soybean

A. Expression cassettes for reducing the proteolytic activity of soybean vacuolar processing enzymes

Soybean plants with reduced vacuolar processing enzyme expression in seed were produced by transformation of plants with expression cassettes designed to knock down expression of the endogenous VPE genes in seed. Two different expression cassettes were each designed and used to independently accomplish this task, one cassette utilized an hpRNA construct in which DNA fragments corresponding to the sequence of the endogenous VPE genes being suppressed is cloned in a loop between two complementary DNA sequences (EL hpRNA; see WO 0200904). The second cassette consisted of an intron-spliced self-complimentary hairpin RNAi (ihpRNA) construct (Smith *et al.* (2000) Nature 407:319-320) designed such that final cassette consisted of two identical ihpRNAs each expressed using an independent promoter.

The loop sequence of the EL hpRNA expression cassette was constructed using standard cloning techniques to splice rtPCR-amplified fragments (293-570 base pairs) of

each of the soy VPE genes (Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a) together in the same sense orientation. The EL hpRNA cassette was then constructed by linking the Kuntz trypsin inhibitor (KTI) promoter (nucleotides 5-2086 of NCBI Accession No. AF233296) the EL DNA sequence, the loop sequence of VPE genes in sense orientation, the EL DNA sequence in reverse orientation (complementary), and the KTI transcriptional termination sequence (nucleotides 2740-2927 of NCBI Accession No. AF233296). SEQ ID NO:15 shows the sequence of this expression cassette.

The stem sequence of the ihpRNA expression cassette was constructed using standard cloning techniques to splice rtPCR-amplified fragments of each of the soy VPE genes (Vpe1a, Vpe1b, Vpe2a, Vpe2b, Vpe3a, and Vpe3b) together. One transcriptional unit of the ihpRNA cassette was then constructed by linking the KTI promoter with the stem sequence fragment in the sense orientation, a PCR-amplified FAD2 intron sequence (nucleotides 142-1274 of NCBI Accession No. AC069473), and the same stem sequence fragment in reverse orientation. The second transcriptional unit of the ihpRNA cassette was constructed in the same fashion with the exception that the late seed preferred (LSP) promoter is substituted for the KTI promoter. The completed ihpRNA expression cassette contained both of these transcriptional units.

B. Transformation

Soybean embryos are then be transformed with the expression cassettes described. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos that produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos that multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 ml liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 ml of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) Nature (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a transgene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al.(1985) Nature 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from E. coli; Gritz et al.(1983) Gene 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium* tumefaciens. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the RNA suppression molecule and or the polypeptide of interest and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL CaCl2 (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds, and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µl of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50

mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

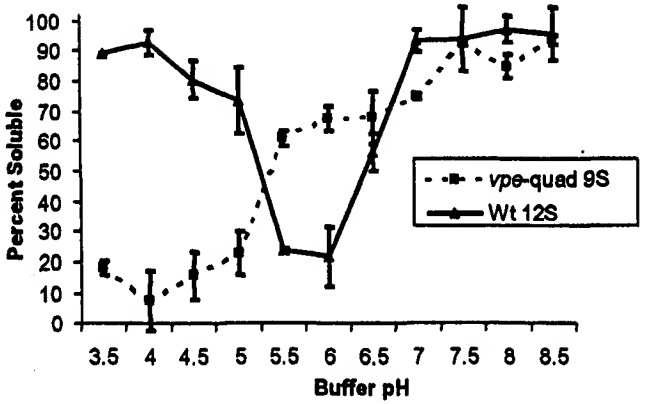
THAT WHICH IS CLAIMED:

1. A soybean plant that is genetically modified to alter one or more functional properties of one or more seed storage proteins, wherein said soybean plant is genetically modified to reduce or eliminate the activity of one or more vacuolar processing enzymes in its seed.

METHODS AND COMPOSITIONS FOR ALTERING THE FUNCTIONAL PROPERTIES OF SEED STORAGE PROTEINS IN SOYBEAN

ABSTRACT OF THE DISCLOSURE

The present invention provides methods and compositions useful for altering the functional properties of soybean seed storage proteins. It is the novel finding of the present invention that the functional properties of seed storage proteins can be altered by reducing the expression of one or more vacuolar processing enzymes in plant seed. Accordingly, in one embodiment, the invention provides a method for altering the functional properties of one or more soybean seed storage proteins. The method comprises transforming a soybean plant cell with at least one expression cassette capable of expressing a polynucleotide that reduces the activity of a vacuolar processing enzyme in the seed of said soybean plant, regenerating a transformed plant from the transformed plant cell, and collecting seed from the regenerated transformed plant. Plants that are genetically modified or mutagenized to alter the functional properties of one or more seed storage proteins, and the transgenic seed of such plants are also provided.



Solubility Properties of Legumin-Type Globulin Protein Isolated from Mature Wild-Type and *vpe*-quad Arabidopsis Seeds

Legumin-type globulin was isolated from sucrose density gradient 12S fractions of wild-type (Wt) gradients and from 9S fractions of *vpe*-quad gradients. Solubility of protein obtained from these fractions was determined under low ionic strength conditions at various pH. Following incubation of the protein sample at a given pH, the amount of protein remaining in solution was quantified and graphed as a percent of the total protein added to the reaction. The error bars show standard deviations (3 replications) at each data point.

FIG. 1

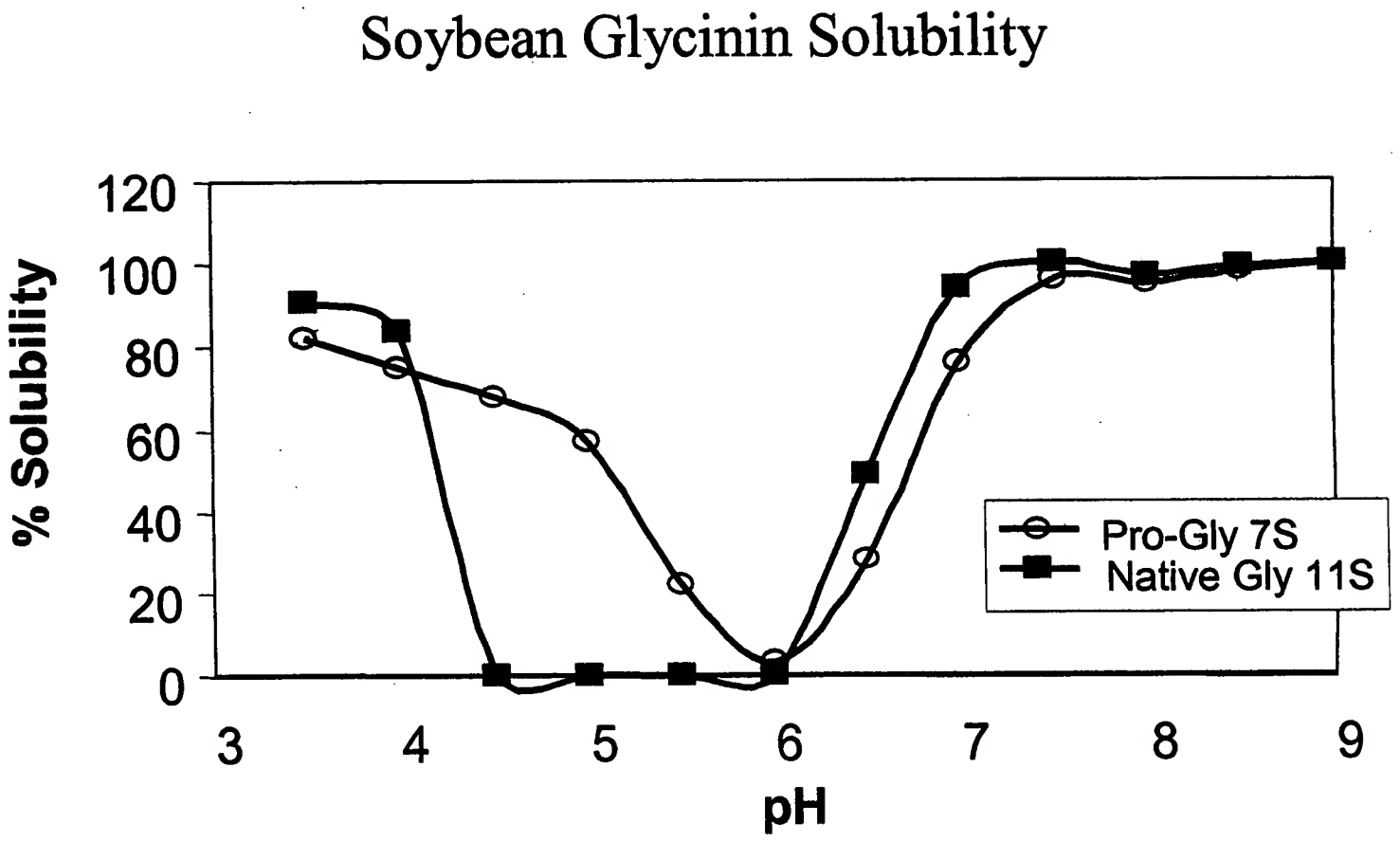

Soybean Glycinin Solubility
% Solubility
120
100
80
60
40
20
0
3
4
5
6
7
8
9
pH
Pro-Gly 7S
Native Gly 11S

FIG. 2